

WO02079447

Publication Title:

AVIAN LYSOZYME PROMOTER

Abstract:

Abstract of WO02079447

The present invention provides novel isolated nucleic acids comprising an avian nucleic acid sequence encoding a lysozyme gene expression control region. The isolated nucleic acid of the present invention is useful for reducing the chromosomal positional effect of a transgene operably linked to the lysozyme gene expression control region and transfected into a recipient cell and allows expression of an operably linked heterologous nucleic acid insert in a transfected avian cells such as, for example, an ov 1007 iduct cell. The isolated avian lysozyme of the present invention may be operably linked with a selected nucleic acid insert encoding a polypeptide desired to be expressed in a transfected cell. The recombinant DNA of the present invention may further comprise a polyadenylation signal sequence or a chicken lysozyme 3' domain.

Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 October 2002 (10.10.2002)

PCT

(10) International Publication Number
WO 02/079447 A2

(51) International Patent Classification⁷: **C12N**
(21) International Application Number: PCT/US02/09866
(22) International Filing Date: 29 March 2002 (29.03.2002)
(25) Filing Language: English
(26) Publication Language: English
(30) Priority Data:
60/280,004 30 March 2001 (30.03.2001) US
09/922,549 3 August 2001 (03.08.2001) US
60/351,550 25 January 2002 (25.01.2002) US

(71) Applicant (for all designated States except US): **AVIGEN-ICS, INC.** [US/US]; 111 Riverbend Road, Athens, GA 30605 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): **RAPP, Jeffrey, C.** [US/US]; 265 Pinewood Circle, Athens, GA 30606 (US).

(74) Agent: **JARECKI-BLACK, Judy, C.**; 111 Riverbend Road, Athens, GA 30605 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

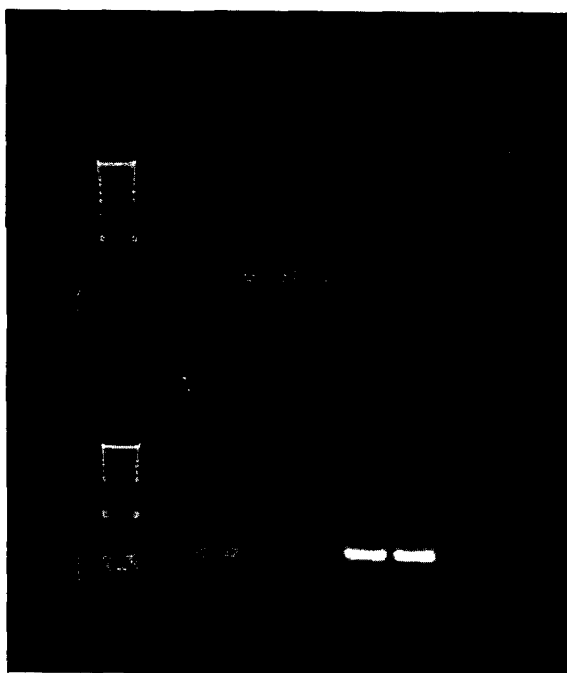
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **AVIAN LYSOZYME PROMOTER**



(57) Abstract: The present invention provides novel isolated nucleic acids comprising an avian nucleic acid sequence encoding a lysozyme gene expression control region. The isolated nucleic acid of the present invention is useful for reducing the chromosomal positional effect of a transgene operably linked to the lysozyme gene expression control region and transfected into a recipient cell and allows expression of an operably linked heterologous nucleic acid insert in a transfected avian cells such as, for example, an oviduct cell. The isolated avian lysozyme of the present invention may be operably linked with a selected nucleic acid insert encoding a polypeptide desired to be expressed in a transfected cell. The recombinant DNA of the present invention may further comprise a polyadenylation signal sequence or a chicken lysozyme 3' domain.



WO 02/079447 A2

Title of the Invention

AVIAN LYSOZYME PROMOTER

The present application is a continuation-in-part of U.S. Serial No. 09/922,549, filed 03 August 2001, and claiming the benefit of priority from provisional application Serial No. 60/280,004, filed March 30, 2001, and provisional application Serial No. 60/351,550 filed January 25, 2002.

Field of the Invention

The present invention relates generally to the identification of an avian lysozyme gene expression control region, specifically from the chicken. More specifically, the invention relates to recombinant nucleic acids and expression vectors, transfected cells and transgenic animals, especially chickens, that comprise the avian lysozyme gene expression control region operably linked to a polypeptide-encoding nucleic acid and, optionally, a chicken lysozyme 3' domain. The present invention further relates to the expression of the polypeptide-encoding nucleic acid under the control of the isolated avian lysozyme gene expression control region.

Background

The field of transgenics was initially developed to understand the action of a single gene in the context of the whole animal and the phenomena of gene activation, expression, and interaction. This technology has also been used to produce models for various diseases in humans and other animals and is amongst the most powerful tools available for the study of genetics, and the understanding of genetic mechanisms and function. From an economic perspective, the use of transgenic technology to convert animals into "protein factories" for the production of specific proteins or other substances of pharmaceutical interest (Gordon et al., 1987, *Biotechnology* 5: 1183-1187; Wilmut et al., 1990, *Theriogenology* 33: 113-123) offers significant advantages over more conventional methods of protein production by gene expression.

Heterologous nucleic acids have been engineered so that an expressed protein may be joined to a protein or peptide that will allow secretion of the transgenic expression product into milk or urine, from which the protein may then be recovered.

These procedures have had limited success and may require lactating animals, with the attendant costs of maintaining individual animals or herds of large species, including cows, sheep, or goats.

Historically, transgenic animals have been produced almost exclusively by microinjection of the fertilized egg. The pronuclei of fertilized eggs are microinjected *in vitro* with foreign, i.e., xenogeneic or allogeneic, heterologous DNA or hybrid DNA molecules. The microinjected fertilized eggs are then transferred to the genital tract of a pseudopregnant female (e.g., Krimpenfort *et al.*, in U.S. Pat. No. 5,175,384).

One system that holds potential is the avian reproductive system. The production of an avian egg begins with formation of a large yolk in the ovary of the hen. The unfertilized oocyte or ovum is positioned on top of the yolk sac. After ovulation, the ovum passes into the infundibulum of the oviduct where it is fertilized, if sperm are present, and then moves into the magnum of the oviduct, which is lined with tubular gland cells. These cells secrete the egg-white proteins, including ovalbumin, lysozyme, ovomucoid, conalbumin and ovomucin, into the lumen of the magnum where they are deposited onto the avian embryo and yolk.

The hen oviduct offers outstanding potential as a protein bioreactor because of the high levels of protein production, the promise of proper folding and post-translation modification of the target protein, the ease of product recovery, and the shorter developmental period of chickens compared to other potential animal species. As a result, efforts have been made to create transgenic chickens expressing heterologous proteins in the oviduct by means of microinjection of DNA (PCT Publication WO 97/47739).

The chicken lysozyme gene is highly expressed in the myeloid lineage of hematopoietic cells, and in the tubular glands of the mature hen oviduct (Hauser *et al.*, 1981, *Hematol. and Blood Transfusion* 26: 175-178; Schutz *et al.*, 1978, Cold Spring Harbor Symp. Quart. Biol. 42: 617-624) and is therefore a suitable candidate for an efficient promoter for heterologous protein production in transgenic animals. The regulatory region of the lysozyme locus extends over at least 12 kb of DNA 5' upstream of the transcription start site, and comprises a number of elements that have

been individually isolated and characterized. The known elements include three enhancer sequences at about -6.1 kb, -3.9 kb, and -2.7 kb (Grewal et al., 1992, *Mol. Cell Biol.* 12: 2339-2350; Banifer et al., 1996, *J. Mol. Med.* 74: 663-671), a hormone responsive element (Hecht et al., 1988, *E.M.B.O.J.* 7: 2063-2073), a silencer element and a complex proximal promoter. The constituent elements of the lysozyme gene expression control region are identifiable as DNAase 1 hypersensitive chromatin sites (DHS). They may be differentially exposed to nuclease digestion depending upon the differentiation stage of the cell. For example, in the multipotent progenitor stage of myelomonocytic cell development, or in erythroblasts, the silencer element is a DHS. At the myeloblast stage, a transcription enhancer located -6.1 kb upstream from the gene transcription start site is a DHS, while at the later monocytic stage another enhancer, at -2.7 kb becomes DNAase sensitive (Huber et al., 1995, *DNA and Cell Biol.* 14: 397-402).

Scattered throughout the chicken genome, including the chicken lysozyme locus, are short stretches of nucleic acid that resemble features of Long Terminal Repeats (LTRs) of retrovirus. The function of these elements is unclear but most likely help define the DHS regions of a gene locus (Stein et al., 1983, *Proc. Natl. Acad. Sci. U.S.A.* 80: 6485-6489).

Flanking the lysozyme gene, including the regulatory region, are matrix attachment regions (5' MAR & 3' MAR), alternatively referred to as "scaffold attachment regions" or SARs. The outer boundaries of the chicken lysozyme locus have been defined by the MARs (Phi-Van et al., 1988, *E.M.B.O.J.* 7: 655-664; Phi-Van, L. and Stratling, W.H., 1996, *Biochem.* 35: 10735-10742). Deletion of a 1.32 kb or a 1.45 kb halves region, each comprising half of a 5' MAR, reduces positional variation in the level of transgene expression (Phi-Van and Stratling, supra).

The 5' matrix-associated region (5' MAR), located about -11.7 kb upstream of the chicken lysozyme transcription start site, can increase the level of gene expression by limiting the positional effects exerted against a transgene (Phi-Van et al., 1988, *supra*). At least one other MAR is located 3' downstream of the protein encoding region. Although MAR nucleic acid sequences are conserved, little cross-

hybridization is seen, indicating significant overall sequence variation. However, MARs of different species can interact with the nucleomatrixes of heterologous species, to the extent that the chicken lysozyme MAR can associate with the plant tobacco nucleomatrix as well as that of the chicken oviduct cells (Mlynarona et al.,
5 1994, *Cell* 6: 417-426; von Kries et al., 1990, *Nucleic Acids Res.* 18: 3881-3885).

Gene expression must be considered not only from the perspective of cis-regulatory elements associated with a gene, and their interactions with trans-acting elements, but also with regard to the genetic environment in which they are located. Chromosomal positioning effects (CPEs), therefore, are the variations in levels of
10 transgene expression associated with different locations of the transgene within the recipient genome. An important factor governing CPE upon the level of transgene expression is the chromatin structure around a transgene, and how it cooperates with the cis-regulatory elements. The cis-elements of the lysozyme locus are confined within a single chromatin domain (Banifer et al., 1996, *supra*; Sippel et al., pgs. 133-
15 147 in Eckstein F. & Lilley D.M.J. (eds), "Nucleic Acids and Molecular Biology", Vol. 3, 1989, Springer).

Deletion of a cis-regulatory element from a transgenic lysozyme locus is sufficient to reduce or eliminate positional independence of the level of gene expression (Banifer et al., 1996, *supra*). There is also evidence indicating that
20 positional independence conferred on a transgene requires the cotransfer of many kilobases of DNA other than just the protein encoding region and the immediate cis-regulatory elements.

The lysozyme promoter region of chicken is active when transfected into mouse fibroblast cells and linked to a reporter gene such as the bacterial
25 chloramphenicol acetyltransferase (CAT) gene. The promoter element is also effective when transiently transfected into chicken promacrophage cells. In each case, however, the presence of a 5' MAR element increased positional independency of the level of transcription (Stief et al., 1989, *Nature* 341: 343-345; Sippel et al., pgs. 257 –
265 in Houdeline L.M. (ed), "Transgenic Animals: Generation and Use").

The ability to direct the insertion of a transgene into a site in the genome of an animal where the positional effect is limited offers predictability of results during the development of a desired transgenic animal, and increased yields of the expressed product. Sippel and Steif disclose, in U.S. Patent No. 5,731,178, methods to increase
5 the expression of genes introduced into eukaryotic cells by flanking a transcription unit with scaffold attachment elements, in particular the 5' MAR isolated from the chicken lysozyme gene. The transcription unit disclosed by Sippel and Steif was an artificial construct that combined only the -6.1 kb enhancer element and the proximal promoter element (base position -579 to +15) from the lysozyme gene. Other
10 promoter associated elements were not included. However, although individual cis-regulatory elements have been isolated and sequenced, together with short regions flanking DNA, the entire nucleic acid sequence comprising the functional 5' upstream region of the lysozyme gene has not been determined in its entirety and therefore not employed as a functional promoter to allow expression of a heterologous transgene.

15 What is still needed, however, is an efficient transcription promoter that will allow expression of a transgene in avian cells that is not subject to positional variation.

What is also needed is a gene expression promoter cassette that will allow expression of a transgene in the oviduct cells of an avian and efficient gene expression
20 regardless of the chromosomal location of the expression system.

Summary of the Invention

Briefly described, the present invention relates to a novel isolated avian nucleic acid comprising an avian lysozyme gene expression control region.

The isolated nucleic acid of the present invention is useful for reducing the
25 chromosomal positional effect of a transgene operably linked to the lysozyme gene expression control region and transfected into a recipient cell. By isolating a region of the avian genome extending from 5' upstream of a 5' MAR of the lysozyme locus to the junction between the signal peptide sequence and a polypeptide-encoding region, cis-elements are also included to allow gene expression in a tissue-specific manner.
30 The lysozyme promoter region of the present invention, therefore, will allow

expression of an operably linked heterologous nucleic acid insert in a transfected avian cell such as, for example, an oviduct cell.

One aspect of the present invention provides a novel isolated nucleic acid that is located immediately 5' upstream of the native lysozyme-encoding region of the chicken lysozyme gene locus. The novel isolated avian nucleic acid sequence encoding a lysozyme gene expression control region comprises at least one 5' matrix attachment region, an intrinsically curved DNA region, at least one transcription enhancer element, a negative regulatory element, at least one hormone responsive element, at least one avian CR1 repeat element, and a proximal lysozyme promoter and signal peptide-encoding region. Interspersed between these constituent elements are stretches of nucleic acid that serve at least to organize the above elements in an ordered array relative to a polypeptide-encoding region.

In one embodiment of the present invention the isolated nucleic acid is isolated from a chicken.

The isolated avian lysozyme of the present invention may be operably linked with a selected nucleic acid insert, wherein the nucleic acid insert encodes a polypeptide desired to be expressed in a transfected cell. The nucleic acid insert may be placed in frame with a signal peptide sequence. Translation initiation may start with the signal peptide and continue through the nucleic acid insert, thereby producing an expressed polypeptide having the desired amino acid sequence.

The sequence of the expressed nucleic acid insert may be optimized for codon usage by a host cell. This may be determined from the codon usage of at least one, and preferably more than one, protein expressed in a chicken cell. For example, the codon usage may be determined from the nucleic acid sequences encoding the proteins ovalbumin, lysozyme, ovomucin and ovotransferrin of chicken.

The recombinant DNA of the present invention may further comprise a polyadenylation signal sequence that will allow the transcript directed by the novel lysozyme gene expression control region to proceed beyond the nucleic acid insert encoding a polypeptide and allow the transcript to further comprise a 3' untranslated region and a polyadenylated tail. Any functional polyadenylation signal sequence may

be linked to the 3' end of the nucleic acid insert including the SV40 polyadenylation signal sequence, bovine growth hormone adenylation sequence or the like.

The recombinant DNA of the present invention may also further comprise the chicken lysozyme 3' domain operably linked to the nucleic acid insert encoding a polypeptide. The 3' domain may include a 3' untranslated region, a polyadenylation
5 signal and a 3' MAR that may reduce positional variation in transgenic avians.

Yet another aspect of the present invention are expression vectors suitable for delivery to a recipient cell for expression of the vector therein. The expression vector of the present invention may comprise an isolated avian lysozyme gene expression
10 control region operably linked to a nucleic acid insert encoding a polypeptide, and optionally a polyadenylation signal sequence. The expression vector may further comprise a bacterial plasmid sequence, a viral nucleic acid sequence, or fragments or variants thereof that may allow for replication of the vector in a suitable host.

Another aspect of the present invention is a method of expressing a heterologous polypeptide in a eukaryotic cell by transfecting the cell with a recombinant DNA comprising an avian lysozyme gene expression control region operably linked to a nucleic acid insert encoding a polypeptide and, optionally, a polyadenylation signal sequence, and culturing the transfected cell in a medium suitable for expression of the heterologous polypeptide under the control of the avian
15 lysozyme gene expression control region.

Also within the scope of the present invention are recombinant cells, tissues and animals containing non-naturally occurring recombinant nucleic acid molecules according to the present invention and described above. In one embodiment of the present invention, the transformed cell is a chicken oviduct cell and the nucleic acid
20 insert comprises the chicken lysozyme gene expression control region, a nucleic acid insert encoding a human interferon $\alpha 2b$ and codon optimized for expression in an avian cell, and an SV40 polyadenylation sequence.

Additional objects and aspects of the present invention will become more apparent upon review of the detailed description set forth below when taken in conjunction with the accompanying figures, which are briefly described as follows.
30

Brief Description of the Figures

Fig. 1 illustrates the primers (SEQ ID NOS: 1-64) used in the sequencing of the lysozyme gene expression control region (SEQ ID NO: 67).

Fig. 2 schematically illustrates the approximately 12 kb lysozyme gene expression control region (SEQ ID NO: 67), indicating the relative positions and orientations of the primers (SEQ ID NOS: 1-64) used in the sequencing thereof.

Fig. 3 illustrates the nucleic acid sequence (SEQ ID NO: 65) comprising the chicken lysozyme gene expression control region (SEQ ID NO: 67), the nucleic acid sequence SEQ ID NO: 66 encoding the chicken expression optimized human interferon α 2b (IFNMAGMAX) and the SV40 polyadenylation signal sequence (SEQ ID NO: 68).

Fig. 4 illustrates the nucleic acid sequence SEQ ID NO: 66 encoding the chicken expression optimized human interferon α 2b (IFNMAGMAX).

Fig. 5 illustrates the nucleic acid sequence SEQ ID NO: 67 encoding the chicken lysozyme gene expression control region.

Fig. 6 illustrates the nucleic acid sequence SEQ ID NO: 68 encoding the SV40 polyadenylation signal sequence.

Fig. 7 illustrates the nucleic acid sequence SEQ ID NO: 69 encoding the chicken lysozyme 3' domain.

Fig. 8 illustrates the nucleic acid sequence SEQ ID NO: 69 encoding the lysozyme gene expression control region (SEQ ID NO: 67) linked to the nucleic acid insert SEQ ID NO: 66 encoding the chicken expression-optimized human interferon α 2b (IFNMAGMAX) and the chicken lysozyme 3' domain SEQ ID NO: 69.

Fig. 9 illustrates the yield of the human interferon α 2b, optimized for chicken expression (IFNMAGMAX), in transfected quail oviduct cultured cells.

Fig 10 illustrates the yield of the human interferon α 2b, optimized for chicken expression (IFNMAGMAX), in chicken myelomonocytic HD11 cells transfected with plasmids pAVIJC-R-A115.93.1.2, pAVIJC-A212.89.2.3 or pAVIJC-R-A212.89.2.1.

Fig 11 illustrates the expression of α 2b human interferon in the blood of transgenic chickens #8305 and #AA61, as compared to standards.

Fig 12 illustrates the gel analysis of PCR products derived from the serum of transgenic birds. Lane and Samples applied to the gel were: 1, marker; 2, 8301; 3, 8303; 4, 8305 5, 8305; 6, 8307; 7, 8309; 8, 8311; 9, marker; 10, 8313; 11, 8305; 12, 8305; 13, Neg. Ctrl; 14, Pos. Ctrl (500pg)+Neg. Ctrl; 15, Pos. Ctrl (500pg).

5 **Detailed Description of the Preferred Embodiments**

Reference now will be made in detail to the presently preferred embodiments of the invention, one or more examples of which are illustrated in the accompanying drawings. Each example is provided by way of explanation of the invention, not limitation of the invention. In fact, it will be apparent to those skilled in the art that
10 various modifications, combinations, additions, deletions and variations can be made in the present invention without departing from the scope or spirit of the invention. For instance, features illustrated or described as part of one embodiment can be used in another embodiment to yield a still further embodiment. It is intended that the present invention covers such modifications, combinations, additions, deletions and
15 variations as come within the scope of the appended claims and their equivalents.

This description uses gene nomenclature accepted by the Cucurbit Genetics Cooperative as it appears in the *Cucurbit Genetics Cooperative Report* 18:85 (1995), herein incorporated by reference in its entirety. Using this gene nomenclature, genes are symbolized by italicized Roman letters. If a mutant gene is recessive to the
20 normal type, then the symbol and name of the mutant gene appear in italicized lower case letters.

For convenience, certain terms employed in the specification, examples, and appended claims are collected here.

Definitions

25 The term "animal" is used herein to include all vertebrate animals, including avians and humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages.

The term "avian" as used herein refers to any species, subspecies or race of organism of the taxonomic class *ava*, such as, but not limited to, such organisms as
30 chicken, turkey, duck, goose, quail, pheasants, parrots, finches, hawks, crows and

ratites including ostrich, emu and cassowary. The term includes the various known strains of *Gallus gallus*, or chickens, (for example, White Leghorn, Brown Leghorn, Barred-Rock, Sussex, New Hampshire, Rhode Island, Australorp, Minorca, Amrox, California Gray, Italian Partidge-colored), as well as strains of turkeys, pheasants, quails, duck, ostriches and other poultry commonly bred in commercial quantities.

The term "nucleic acid" as used herein refers to any natural and synthetic linear and sequential arrays of nucleotides and nucleosides, for example cDNA, genomic DNA, mRNA, tRNA, oligonucleotides, oligonucleosides and derivatives thereof. For ease of discussion, such nucleic acids may be collectively referred to herein as "constructs," "plasmids," or "vectors." Representative examples of the nucleic acids of the present invention include bacterial plasmid vectors including expression, cloning, cosmid and transformation vectors such as, but not limited to, pBR322, animal viral vectors such as, but not limited to, modified adenovirus, influenza virus, polio virus, pox virus, retrovirus, and the like, vectors derived from bacteriophage nucleic acid, and synthetic oligonucleotides like chemically synthesized DNA or RNA. The term "nucleic acid" further includes modified or derivatised nucleotides and nucleosides such as, but not limited to, halogenated nucleotides such as, but not only, 5-bromouracil, and derivatised nucleotides such as biotin-labeled nucleotides.

The term "isolated nucleic acid" as used herein refers to a nucleic acid with a structure (a) not identical to that of any naturally occurring nucleic acid or (b) not identical to that of any fragment of a naturally occurring genomic nucleic acid spanning more than three separate genes, and includes DNA, RNA, or derivatives or variants thereof. The term covers, for example, (a) a DNA which has the sequence of part of a naturally occurring genomic molecule but is not flanked by at least one of the coding sequences that flank that part of the molecule in the genome of the species in which it naturally occurs; (b) a nucleic acid incorporated into a vector or into the genomic nucleic acid of a prokaryote or eukaryote in a manner such that the resulting molecule is not identical to any vector or naturally occurring genomic DNA; (c) a separate molecule such as a cDNA, a genomic fragment, a fragment produced by

polymerase chain reaction (PCR), ligase chain reaction (LCR) or chemical synthesis, or a restriction fragment; (d) a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein, and (e) a recombinant nucleotide sequence that is part of a hybrid sequence that is not naturally occurring. Isolated
5 nucleic acid molecules of the present invention can include, for example, natural allelic variants as well as nucleic acid molecules modified by nucleotide deletions, insertions, inversions, or substitutions such that the resulting nucleic acid molecule still essentially encodes a lysozyme gene expression control region or a variant thereof of the present invention.

10 By the use of the term “enriched” in reference to nucleic acid it is meant that the specific DNA or RNA sequence constitutes a significantly higher fraction of the total DNA or RNA present in the cells or solution of interest than in normal or diseased cells or in the cells from which the sequence was taken. Enriched does not imply that there are no other DNA or RNA sequences present, just that the relative
15 amount of the sequence of interest has been significantly increased. The other DNA may, for example, be derived from a yeast or bacterial genome, or a cloning vector, such as a plasmid or a viral vector. The term “significant” as used herein is used to indicate that the level of increase is useful to the person making such an increase.

It is advantageous for some purposes that a nucleotide sequence is in purified
20 form. The term “purified” in reference to nucleic acid represents that the sequence has increased purity relative to the natural environment.

The terms “polynucleotide,” “oligonucleotide,” and “nucleic acid sequence” are used interchangeably herein and include, but are not limited to, coding sequences (polynucleotide(s) or nucleic acid sequence(s) which are transcribed and translated
25 into polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory or control sequences); control sequences (e.g., translational start and stop codons, promoter sequences, ribosome binding sites, polyadenylation signals, transcription factor binding sites, transcription termination sequences, upstream and downstream regulatory domains, enhancers, silencers, and the like); and regulatory
30 sequences (DNA sequences to which a transcription factor(s) binds and alters the

activity of a gene's promoter either positively (induction) or negatively (repression)). No limitation as to length or to synthetic origin are suggested by the terms described herein.

As used herein the terms "polypeptide" and "protein" refer to a polymer of amino acids of three or more amino acids in a serial array, linked through peptide bonds. The term "polypeptide" includes proteins, protein fragments, protein analogues, oligopeptides and the like. The term "polypeptides" contemplates polypeptides as defined above that are encoded by nucleic acids, produced through recombinant technology (isolated from an appropriate source such as a bird), or synthesized. The term "polypeptides" further contemplates polypeptides as defined above that include chemically modified amino acids or amino acids covalently or noncovalently linked to labeling ligands.

The term "fragment" as used herein to refer to a nucleic acid (e.g., cDNA) refers to an isolated portion of the subject nucleic acid constructed artificially (e.g., by chemical synthesis) or by cleaving a natural product into multiple pieces, using restriction endonucleases or mechanical shearing, or a portion of a nucleic acid synthesized by PCR, DNA polymerase or any other polymerizing technique well known in the art, or expressed in a host cell by recombinant nucleic acid technology well known to one of skill in the art. The term "fragment" as used herein may also refer to an isolated portion of a polypeptide, wherein the portion of the polypeptide is cleaved from a naturally occurring polypeptide by proteolytic cleavage by at least one protease, or is a portion of the naturally occurring polypeptide synthesized by chemical methods well known to one of skill in the art.

The term "gene" or "genes" as used herein refers to nucleic acid sequences (including both RNA or DNA) that encode genetic information for the synthesis of a whole RNA, a whole protein, or any portion of such whole RNA or whole protein. Genes that are not naturally part of a particular organism's genome are referred to as "foreign genes," "heterologous genes" or "exogenous genes" and genes that are naturally a part of a particular organism's genome are referred to as "endogenous genes". The term "gene product" refers to RNAs or proteins that are encoded by the

gene. "Foreign gene products" are RNA or proteins encoded by "foreign genes" and
"endogenous gene products" are RNA or proteins encoded by endogenous genes.
"Heterologous gene products" are RNAs or proteins encoded by "foreign,
heterologous or exogenous genes" and are, therefore, not naturally expressed in the
5 cell.

The term "expressed" or "expression" as used herein refers to the transcription
from a gene to give an RNA nucleic acid molecule at least complementary in part to a
region of one of the two nucleic acid strands of the gene. The term "expressed" or
"expression" as used herein also refers to the translation from said RNA nucleic acid
10 molecule to give a protein, a polypeptide or a portion thereof.

As used herein, the term "locus" or "loci" refers to the site of a gene on a
chromosome. Pairs of genes control hereditary traits, each in the same position on a
pair of chromosomes. These gene pairs, or alleles, may both be dominant or both be
recessive in expression of that trait. In either case, the individual is said to be
15 homozygous for the trait controlled by that gene pair. If the gene pair (alleles)
consists of one dominant and one recessive trait, the individual is heterozygous for the
trait controlled by the gene pair. Natural variation in genes or nucleic acid molecules
caused by, for example, recombination events or resulting from mutation, gives rise to
allelic variants with similar, but not identical, nucleotide sequences. Such allelic
20 variants typically encode proteins with similar activity to that of the protein encoded
by the gene to which they are compared, because natural selection typically selects
against variations that alter function. Allelic variants can also comprise alterations in
the untranslated regions of the gene as, for example, in the 3' or 5' untranslated regions
or can involve alternate splicing of a nascent transcript, resulting in alternative exons
25 being positioned adjacently.

The term "operably linked" refers to an arrangement of elements wherein the
components so described are configured so as to perform their usual function. Control
sequences operably linked to a coding sequence are capable of effecting the
expression of the coding sequence. The control sequences need not be contiguous with
30 the coding sequence, so long as they function to direct the expression thereof. Thus,

for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered “operably linked” to the coding sequence.

The terms “transcription regulatory sequences” and “gene expression control regions” as used herein refer to nucleotide sequences that are associated with a gene
5 nucleic acid sequence and which regulate the transcriptional expression of the gene. Exemplary transcription regulatory sequences include enhancer elements, hormone response elements, steroid response elements, negative regulatory elements, and the like. The “transcription regulatory sequences” may be isolated and incorporated into a
10 vector nucleic acid to enable regulated transcription in appropriate cells of portions of the vector DNA. The “transcription regulatory sequence” may precede, but is not limited to, the region of a nucleic acid sequence that is in the region 5' of the end of a protein coding sequence that may be transcribed into mRNA. Transcriptional regulatory sequences may also be located within a protein coding region, in regions of
15 a gene that are identified as “intron” regions, or may be in regions of nucleic acid sequence that are in the region of nucleic acid.

The term “promoter” as used herein refers to the DNA sequence that determines the site of transcription initiation from an RNA polymerase. A “promoter-proximal element” may be a regulatory sequence within about 200 base pairs of the
20 transcription start site.

The terms “matrix attachment regions” or “SAR elements” as used herein refer to DNA sequences having an affinity or intrinsic binding ability for the nuclear scaffold or matrix. The MAR elements of the chicken lysozyme locus were described by Phi-Van *et al.*, 1988, *E.M.B.O. J.* 76: 665-664 and Phi-Van, L. and Stratling, W.H.,
25 1996, *Biochem.* 35: 10735-10742, the contents of which are incorporated herein by reference in their entireties.

The term “coding region” as used herein refers to a continuous linear arrangement of nucleotides which may be translated into a protein. A full length coding region is translated into a full length protein; that is, a complete protein as
30 would be translated in its natural state absent any post-translational modifications. A

full length coding region may also include any leader protein sequence or any other region of the protein that may be excised naturally from the translated protein.

The term "complementary" as used herein refers to two nucleic acid molecules that can form specific interactions with one another. In the specific interactions, an adenine base within one strand of a nucleic acid can form two hydrogen bonds with thymine within a second nucleic acid strand when the two nucleic acid strands are in opposing polarities. Also in the specific interactions, a guanine base within one strand of a nucleic acid can form three hydrogen bonds with cytosine within a second nucleic acid strand when the two nucleic acid strands are in opposing polarities. Complementary nucleic acids as referred to herein, may further comprise modified bases wherein a modified adenine may form hydrogen bonds with a thymine or modified thymine, and a modified cytosine may form hydrogen bonds with a guanine or a modified guanine.

The term "probe" as used herein, when referring to a nucleic acid, refers to a nucleotide sequence that can be used to hybridize with and thereby identify the presence of a complementary sequence, or a complementary sequence differing from the probe sequence but not to a degree that prevents hybridization under the hybridization stringency conditions used. The probe may be modified with labels such as, but not only, radioactive groups, biotin, and the like that are well known in the art.

The term "capable of hybridizing under stringent conditions" as used herein refers to annealing a first nucleic acid to a second nucleic acid under stringent conditions as defined below. Stringent hybridization conditions typically permit the hybridization of nucleic acid molecules having at least 70% nucleic acid sequence identity with the nucleic acid molecule being used as a probe in the hybridization reaction. For example, the first nucleic acid may be a test sample or probe, and the second nucleic acid may be the sense or antisense strand of a lysozyme gene expression control region or a fragment thereof. Hybridization of the first and second nucleic acids may be conducted under stringent conditions, e.g., high temperature and/or low salt content that tend to disfavor hybridization of dissimilar nucleotide

sequences. Alternatively, hybridization of the first and second nucleic acid may be conducted under reduced stringency conditions, e.g., low temperature and/or high salt content that tend to favor hybridization of dissimilar nucleotide sequences. Low stringency hybridization conditions may be followed by high stringency conditions or
5 intermediate medium stringency conditions to increase the selectivity of the binding of the first and second nucleic acids. The hybridization conditions may further include reagents such as, but not limited to, dimethyl sulfoxide (DMSO) or formamide to disfavor still further the hybridization of dissimilar nucleotide sequences. A suitable hybridization protocol may, for example, involve hybridization in 6X SSC (wherein
10 1X SSC comprises 0.015 M sodium citrate and 0.15 M sodium chloride), at 65° C in an aqueous solution, followed by washing with 1X SSC at 65° C. Formulae to calculate appropriate hybridization and wash conditions to achieve hybridization permitting 30% or less mismatch between two nucleic acid molecules are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138: 267-284; the content of
15 which is herein incorporated by reference in its entirety. Protocols for hybridization techniques are well known to those of skill in the art and standard molecular biology manuals may be consulted to select a suitable hybridization protocol without undue experimentation. See, for example, Sambrook et al., 1989, "Molecular Cloning: A Laboratory Manual", 2nd ed., Cold Spring Harbor Press, the contents of which are
20 herein incorporated by reference in its entirety.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) from about pH 7.0 to about pH 8.3 and the temperature is at least about 30° C for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C for long
25 probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37° Celsius, and a wash in 1x to 2x SSC at 50 to 55° Celsius. Exemplary moderate stringency conditions
30 include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37° Celsius,

and a wash in 0.5x to 1x SSC at 55 to 60° Celsius. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37° Celsius, and a wash in 0.1x SSC at 60 to 65° Celsius.

The terms “unique nucleic acid region” and “unique protein (polypeptide) region” as used herein refer to sequences present in a nucleic acid or protein (polypeptide) respectively that is not present in any other nucleic acid or protein sequence. The terms “conserved nucleic acid region” as referred to herein is a nucleotide sequence present in two or more nucleic acid sequences, to which a particular nucleic acid sequence can hybridize under low, medium or high stringency conditions. The greater the degree of conservation between the conserved regions of two or more nucleic acid sequences, the higher the hybridization stringency that will allow hybridization between the conserved region and a particular nucleic acid sequence.

The terms “percent sequence identity” or “percent sequence similarity” as used herein refer to the degree of sequence identity between two nucleic acid sequences or two amino acid sequences as determined using the algorithm of Karlin and Attschul, 1990, *Proc. Natl. Acad. Sci.* 87: 2264-2268, modified as in Karlin and Attschul, 1993, *Proc. Natl. Acad. Sci.* 90: 5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Attschul et al., 1990, *T. Mol. Biol.* Q15: 403-410. BLAST nucleotide searches are performed with the NBLAST program, score = 100, wordlength = 12, to obtain nucleotide sequences homologous to a nucleic acid molecule of the invention. BLAST protein searches are performed with the XBLAST program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a reference polypeptide. To obtain gapped alignments for comparison purposes, Gapped BLAST is utilized as described in Attschul et al., 1997, *Nucl. Acids Res.* 25: 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g. XBLAST and NBLAST) are used. See <http://www.ncbi.nlm.nih.gov>. Other algorithms, programs and default settings may also be suitable such as, but not only, the GCG-Sequence Analysis Package of the U.K. Human Genome Mapping Project Resource Centre that includes programs for

nucleotide or amino acid sequence comparisons.

The term “sense strand” as used herein refers to a single stranded DNA molecule from a genomic DNA that may be transcribed into RNA and translated into the natural polypeptide product of the gene. The term “antisense strand” as used
5 herein refers to the single strand DNA molecule of a genomic DNA that is complementary with the sense strand of the gene.

The term “antisense DNA” as used herein refers to a gene sequence DNA that has a nucleotide sequence complementary to the “sense strand” of a gene when read in reverse orientation, i.e., DNA read into RNA in a 3' to 5' direction rather than in the 5'
10 to 3' direction. The term “antisense RNA” is used to mean an RNA nucleotide sequence (for example that encoded by an antisense DNA or synthesized complementary with the antisense DNA). Antisense RNA is capable of hybridizing under stringent conditions with an antisense DNA. The antisense RNA of the invention is useful for regulating expression of a “target gene” either at the
15 transcriptional or translational level. For example, transcription of the subject nucleic acids may produce antisense transcripts that are capable of inhibiting transcription by inhibiting initiation of transcription or by competing for limiting transcription factors; the antisense transcripts may inhibit transport of the “target RNA”, or, the antisense transcripts may inhibit translation of “target RNA”.

20 The term “nucleic acid vector” as used herein refers to a natural or synthetic single or double stranded plasmid or viral nucleic acid molecule that can be transfected or transformed into cells and replicate independently of, or within, the host cell genome. A circular double stranded plasmid can be linearized by treatment with an appropriate restriction enzyme based on the nucleotide sequence of the plasmid
25 vector. A nucleic acid can be inserted into a vector by cutting the vector with restriction enzymes and ligating the pieces together. The nucleic acid molecule can be RNA or DNA.

The term “expression vector” as used herein refers to a nucleic acid vector that comprises the lysozyme gene expression control region operably linked to a
30 nucleotide sequence coding at least one polypeptide. As used herein, the term

“regulatory sequences” includes promoters, enhancers, and other elements that may control gene expression. Standard molecular biology textbooks such as Sambrook et al. eds., 1989, “Molecular Cloning: A Laboratory Manual”, 2nd ed., Cold Spring Harbor Press may be consulted to design suitable expression vectors that may further
5 include an origin of replication and selectable gene markers. It should be recognized, however, that the choice of a suitable expression vector and the combination of functional elements therein depends upon multiple factors including the choice of the host cell to be transformed and/or the type of protein to be expressed.

The terms “transformation” and “transfection” as used herein refer to the
10 process of inserting a nucleic acid into a host. Many techniques are well known to those skilled in the art to facilitate transformation or transfection of a nucleic acid into a prokaryotic or eukaryotic organism. These methods involve a variety of techniques, such as treating the cells with high concentrations of salt such as, but not only a calcium or magnesium salt, an electric field, detergent, or liposome mediated
15 transfection, to render the host cell competent for the uptake of the nucleic acid molecules, and by such methods as sperm-mediated and restriction-mediated integration.

The term “transfecting agent” as used herein refers to a composition of matter added to the genetic material for enhancing the uptake of heterologous DNA
20 segment(s) into a eukaryotic cell, preferably an avian cell, and more preferably a chicken male germ cell. The enhancement is measured relative to the uptake in the absence of the transfecting agent. Examples of transfecting agents include adenovirus-transferrin-polylysine-DNA complexes. These complexes generally augment the uptake of DNA into the cell and reduce its breakdown during its passage
25 through the cytoplasm to the nucleus of the cell. These complexes can be targeted to the male germ cells using specific ligands that are recognized by receptors on the cell surface of the germ cell, such as the c-kit ligand or modifications thereof.

Other preferred transfecting agents include but are not limited to lipofectin, lipfectamine, DIMRIE C, Supeffect, and Effectin (Qiagen), unifectin, maxifectin,
30 DOTMA, DOGS (Transfectam; dioctadecylamidoglycylspermine), DOPE (1,2-

dioleoyl-sn-glycero-3-phosphoethanolamine), DOTAP (1,2-dioleoyl-3-trimethylammonium propane), DDAB (dimethyl dioctadecylammonium bromide), DHDEAB (N,N-di-n-hexadecyl-N,N-dihydroxyethyl ammonium bromide), HDEAB (N-n-hexadecylN,N-dihydroxyethylammonium bromide), polybrene, or
5 poly(ethylenimine) (PEI). These non-viral agents have the advantage that they can facilitate stable integration of xenogeneic DNA sequences into the vertebrate genome, without size restrictions commonly associated with virus-derived transfecting agents.

The term "recombinant cell" refers to a cell that has a new combination of nucleic acid segments that are not covalently linked to each other in nature. A new
10 combination of nucleic acid segments can be introduced into an organism using a wide array of nucleic acid manipulation techniques available to those skilled in the art. A recombinant cell can be a single eukaryotic cell, or a single prokaryotic cell, or a mammalian cell. The recombinant cell may harbor a vector that is extragenomic. An extragenomic nucleic acid vector does not insert into the cell's genome. A
15 recombinant cell may further harbor a vector or a portion thereof that is intragenomic. The term intragenomic defines a nucleic acid construct incorporated within the recombinant cell's genome.

The terms "recombinant nucleic acid" and "recombinant DNA" as used herein refer to combinations of at least two nucleic acid sequences that are not naturally
20 found in a eukaryotic or prokaryotic cell. The nucleic acid sequences may include, but are not limited to, nucleic acid vectors, gene expression regulatory elements, origins of replication, suitable gene sequences that when expressed confer antibiotic resistance, protein-encoding sequences and the like. The term "recombinant polypeptide" is meant to include a polypeptide produced by recombinant DNA
25 techniques such that it is distinct from a naturally occurring polypeptide either in its location, purity or structure. Generally, such a recombinant polypeptide will be present in a cell in an amount different from that normally observed in nature.

Pharmaceutical compositions comprising agents that will modulate the regulation of the expression of a polypeptide-encoding nucleic acid operably linked to
30 a lysozyme gene expression control region can be administered in dosages and by

techniques well known to those skilled in the medical or veterinary arts, taking into consideration such factors as the age, sex, weight, species and condition of the recipient animal, and the route of administration. The route of administration can be percutaneous, via mucosal administration (e.g., oral, nasal, anal, vaginal) or via a
5 parenteral route (intradermal, intramuscular, subcutaneous, intravenous, or intraperitoneal). Pharmaceutical compositions can be administered alone, or can be co-administered or sequentially administered with other treatments or therapies. Forms of administration may include suspensions, syrups or elixirs, and preparations for parenteral, subcutaneous, intradermal, intramuscular or intravenous administration
10 (e.g., injectable administration) such as sterile suspensions or emulsions. Pharmaceutical compositions may be administered in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose, or the like. The compositions can contain auxiliary substances such as wetting or emulsifying agents, pH buffering agents, adjuvants, gelling or viscosity enhancing
15 additives, preservatives, flavoring agents, colors, and the like, depending upon the route of administration and the preparation desired. Standard pharmaceutical texts, such as "Remington's Pharmaceutical Science", 17th edition, 1985 may be consulted to prepare suitable preparations, without undue experimentation. Dosages can generally range from a few hundred milligrams to a few grams.

20 As used herein, a "transgenic animal" is any animal, such as an avian species, including the chicken, in which one or more of the cells of the avian may contain heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into a cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate
25 genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA. In the typical transgenic animal, the transgene
30 causes cells to express a recombinant form of the subject polypeptide, e.g., either

agonistic or antagonistic forms, or in which the gene has been disrupted. The terms “chimeric animal” or “mosaic animal” are used herein to refer to animals in which the recombinant gene is found, or in which the recombinant is expressed in some but not all cells of the animal. The term “tissue-specific chimeric animal” indicates that the recombinant gene is present and/or expressed in some tissues but not others.

As used herein, the term “transgene” means a nucleic acid sequence (encoding, for example, a human interferon polypeptide) that is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the animal's genome in such a way as to alter the genome of the cell into which it is inserted (*e.g.*, it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene according to the present invention will include one or more transcriptional regulatory sequences, polyadenylation signal sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of a selected nucleic acid.

The term “chromosomal positional effect (CPE)” as used herein refers to the variation in the degree of gene transcription as a function of the location of the transcribed locus within the cell genome. Random transgenesis may result in a transgene being inserted at different locations in the genome so that individual cells of a population of transgenic cells may each have at least one transgene, each at a different location and therefore each in a different genetic environment. Each cell, therefore, may express the transgene at a level specific for that particular cell and dependent upon the immediate genetic environment of the transgene. In a transgenic animal, as a consequence, different tissues may exhibit different levels of transgene expression.

Techniques useful for isolating and characterizing the nucleic acids and proteins of the present invention are well known to those of skill in the art and standard molecular biology and biochemical manuals may be consulted to select suitable protocols without undue experimentation. See, for example, Sambrook *et al.*,

1989, "Molecular Cloning: A Laboratory Manual", 2nd ed., Cold Spring Harbor, the content of which is herein incorporated by reference in its entirety.

Abbreviations:

Abbreviations used in the present specification include the following: aa, amino acid(s); bp, base pair(s); cDNA, DNA complementary to RNA; nt, nucleotide(s); SSC, sodium chloride-sodium citrate; DMSO, dimethyl sulfoxide; MAR; matrix attachment region.

Chicken lysozyme gene expression control region nucleic acid sequences: A series of PCR amplifications of template chicken genomic DNA were used to isolate the gene expression control region of the chicken lysozyme locus. Two amplification reactions used the PCR primer sets SEQ ID NOS: 1 and 2 and SEQ ID NOS: 3 and 4. The amplified PCR products were united as a contiguous isolated nucleic acid by a third PCR amplification step with the primers SEQ ID NOS: 1 and 4, as described in Example 1 below.

The isolated PCR-amplified product, comprising about 12 kb of the nucleic acid region 5' upstream of the native chicken lysozyme gene locus, was cloned into the plasmid pCMV-LysSPIFNMM. pCMV-LysSPIFNMM comprises a modified nucleic acid insert encoding a human interferon $\alpha 2b$ sequence and an SV40 polyadenylation signal sequence 3' downstream of the interferon encoding nucleic acid. The sequence SEQ ID NO: 66 of the nucleic acid insert encoding human interferon $\alpha 2b$ was in accordance with avian cell codon usage, as determined from the nucleotide sequences encoding chicken ovomucin, ovalbumin, ovotransferrin and lysozyme. The novel chicken lysozyme gene expression control region, interferon-encoding insert and the SV40 polyadenylation signal sequence of the resulting plasmid construct pAVIJCR-A115.93.1.2, constructed as described in Example 1 below, was sequenced using the artificial oligonucleotide primers SEQ ID NOS: 1-64, as illustrated in Figs. 1 and 2.

The nucleic acid sequence (SEQ ID NO: 65) (GenBank Accession No. AF405538) of the insert in pAVIJCR-A115.93.1.2 is shown in Fig. 3, with the modified human interferon $\alpha 2b$ encoding nucleotide sequence SEQ ID NO: 66

(GenBank Accession No. AF405539) and the novel chicken lysozyme gene expression control region SEQ ID NO: 67 (GenBank Accession No. AF405540) shown in Figs. 4 and 5 respectively. A polyadenylation signal sequence that is suitable for operably linking to the polypeptide-encoding nucleic acid insert is the SV40 signal sequence SEQ ID NO: 68, as shown in Fig. 6.

The plasmid pAVIJCR-A115.93.1.2 was restriction digested with enzyme *FseI* to isolate a 15.4 kb DNA containing the lysozyme 5' matrix attachment region (MAR) and the -12.0 kb lysozyme promoter during the expression of the interferon-encoding insert, as described in Example 2, below. Plasmid pIIIilys was restriction digested with *MluI* and *XhoI* to isolate an approximately 6 kb nucleic acids, comprising the 3' lysozyme domain, the sequence of which (SEQ ID NO: 70) is shown in Fig. 7. The 15.4 kb and 6 kb nucleic acids were ligated and the 21.4 kb nucleic acid comprising the nucleic acid sequence SEQ ID NO: 70 (GenBank Assession No. AF 497473) as shown in Fig. 8 was transformed into recipient STBL4 cells as described in Example 2, below.

The inclusion of the novel isolated avian lysozyme gene expression control region of the present invention upstream of a codon-optimized interferon-encoding sequence in pAVIJCR-A115.93.1.2 allowed expression of the interferon polypeptide in transfected avian cells, as described in Example 5, below. The 3' lysozyme domain SEQ ID NO: 69, when operably linked downstream of the heterologous nucleic acid insert, also allows expression of the nucleic acid insert as described in Example 7, below. For example, the nucleic acid insert may encode a heterologous polypeptide such as the $\alpha 2b$ interferon having sequence SEQ ID NO: 66.

It is further contemplated that any nucleic acid sequence encoding a polypeptide may be operably linked to the novel isolated avian lysozyme gene expression control region and optionally operably linked to the 3' lysozyme domain SEQ ID NO. 69 so as to be expressed in a transfected avian cell. The plasmid construct pAVIJCR-A115.93.1.2 was transfected into cultured quail oviduct cells, which were then incubated for about 72 hours. ELISA assays of the cultured media showed that the transfected cells synthesized a polypeptide detectable with anti-human

interferon α 2b antibodies. Plasmid construct pAVIJCR-A212.89.2.1 and pAVIJCR-A212.89.2.3 transfected into chicken myelomonocytic HD11 cells yield detectable human α 2b interferon, as described in Example 9 and shown in Fig. 10, below.

5 The novel isolated chicken lysozyme gene expression control region of the present invention comprises the nucleotide elements that are positioned 5' upstream of the lysozyme-encoding region of the native chicken lysozyme locus and which are necessary for the regulated expression of a downstream polypeptide-encoding nucleic acid. While not wishing to be bound by any one theory, the inclusion of at least one 5' MAR element in the isolated control region may confer positional independence to
10 a transfected gene operably linked to the novel lysozyme gene expression control region.

The isolated lysozyme gene expression control region of the present invention is useful for reducing the chromosomal positional effect of a transgene operably linked to the lysozyme gene expression control region and transfected into a recipient
15 avian cell. By isolating a region of the avian genome extending from a point 5' upstream of a 5' MAR of the lysozyme locus to the junction between the signal peptide sequence and a polypeptide-encoding region, cis-regulatory elements are also included that may allow gene expression in a tissue-specific manner. The lysozyme promoter region of the present invention, therefore, will allow expression of an
20 operably linked heterologous nucleic acid insert in a transfected avian cell such as, for example, an oviduct cell.

It is further contemplated that a recombinant DNA of the present invention may further comprise the chicken lysozyme 3' domain (SEQ. ID NO: 69) linked downstream of the nucleic acid insert encoding a heterologous polypeptide. The
25 lysozyme 3' domain includes a nucleic acid sequence encoding a 3' MAR domain that may cooperate with a 5' MAR to direct the insertion of the construct of the present invention into the chromosome of a transgenic avian, or may act independently of the 5' MAR.

One aspect of the present invention, therefore, provides a novel isolated
30 nucleic acid that comprises the nucleotide sequence SEQ ID NO: 67, shown in Fig. 5

and derivatives and variants thereof, that is located immediately 5' upstream of the native lysozyme-encoding region of the chicken lysozyme gene locus.

In one embodiment of the novel isolated nucleic acid of the present invention, therefore, the avian nucleic acid sequence encoding a lysozyme gene expression
5 control region comprises at least one 5' matrix attachment region, an intrinsically curved DNA region, at least one transcription enhancer element, a negative regulatory element, at least one hormone responsive element, at least one avian CR1 repeat element, and a proximal lysozyme promoter and signal peptide-encoding region. Interspersed between these constituent elements are stretches of nucleic acid that serve
10 at least to organize the above elements in an ordered array relative to a polypeptide-encoding region, such as that encoding for chicken lysozyme. It is contemplated to be within the scope of the present invention that the cis-elements of the lysozyme gene expression control region may be in any linear arrangement that can allow the formation of a transcript comprising the nucleotide sequence or its complement of a
15 nucleic insert operably linked to the lysozyme gene expression control region.

In one embodiment of the present invention, the isolated nucleic acid may be isolated from an avian selected from the group consisting of a chicken, a turkey, a duck, a goose, a quail, a pheasant, a ratite, an ornamental bird or a feral bird.

In another embodiment of the present invention, the isolated nucleic acid is
20 obtained from a chicken. In this embodiment, the isolated nucleic acid has the sequence of SEQ ID NO: 67, as shown in Fig. 5, or a variant thereof.

Another aspect of the present invention provides a novel isolated nucleic acid that comprises the nucleic acid SEQ ID NO: 69 encoding the chicken 3' lysozyme domain operably linked to the nucleic acid having sequence SEQ ID NO: 65.

25 One embodiment of the isolated nucleic acid of the present invention, therefore, is a lysozyme gene expression control region comprising the nucleic acid sequence SEQ ID NO. 66 operably linked to a nucleic acid for expression in avian cells, and a chicken 3' lysozyme domain having the nucleic acid SEQ ID NO: 70, as shown in Fig. 8.

In another embodiment of the isolated nucleic acid of the present invention, the nucleic acid for expression in avian cells encodes a human interferon $\alpha 2b$ protein optimized for expression in avian cells.

Another aspect of the invention provides nucleic acids that can hybridize under
5 high, medium or low stringency conditions to an isolated nucleic acid that encodes a chicken lysozyme gene expression control region having all, a derivative of, or a portion of the nucleic acid sequence SEQ ID NO: 67 shown in Fig. 5. The nucleotide sequence determined from the isolation of the lysozyme gene expression control region from a chicken (SEQ ID NO: 67) will allow for the generation of probes
10 designed for use in identifying homologs of lysozyme gene expression control regions in other avian species.

Fragments of a nucleic acid encoding a portion of the subject lysozyme gene expression control region are also within the scope of the invention. As used herein, a fragment of the nucleic acid encoding an active portion of a lysozyme gene expression
15 control region refers to a nucleotide sequence having fewer nucleotides than the nucleotide sequence encoding the entire nucleic acid sequence of the lysozyme gene expression control region.

In one embodiment of the present invention, the nucleotide sequence of the isolated DNA molecule of the present invention may be used as a probe in nucleic
20 acid hybridization assays for the detection of the lysozyme gene expression control region. The nucleotide sequence of the present invention may be used in any nucleic acid hybridization assay system known in the art, including, but not limited to, Southern blots (Southern, E.M., 1975, *J. Mol. Biol.* 98: 508), Northern blots (Thomas et al., 1980, *Proc. Natl. Acad. Sci.* 77: 5201-05), and Colony blots (Grunstein et al.,
25 1975, *Proc. Natl. Acad. Sci.* 72: 3961-65)(the contents of which are hereby incorporated by reference in their entireties). Alternatively, the isolated DNA molecules of the present invention can be used in a gene amplification detection procedure such as a polymerase chain reaction (Erich et al., 1991, *Science* 252: 1643-51, the content of which is hereby incorporated by reference in its entirety) or in
30 restriction fragment length polymorphism (RFLP) diagnostic techniques, as described

in pgs. 519-522 and 545-547 of Watson *et al.*, 2nd ed., 1992, "Recombinant DNA", Scientific American Books (the contents of which is hereby incorporated by reference in its entirety).

5 Nucleotides constructed in accordance with the present invention can be labeled to provide a signal as a means of detection. For example, radioactive elements such as ^{32}P , ^3H , and ^{35}S or the like provide sufficient half-life to be useful as radioactive labels. Other materials useful for labeling synthetic nucleotides include fluorescent compounds, enzymes and chemiluminescent moieties. Methods useful in selecting appropriate labels and binding protocols for binding the labels to the
10 synthetic nucleotides are well known to those of skill in the art. Standard immunology manuals, such as Promega: Protocol and Applications Guide, 2nd Edition, 1991 (Promega Corp., Madison, WI, the content of which is incorporated herein in its entirety), may be consulted to select an appropriate labeling protocol without undue experimentation.

15 In another embodiment of the present invention, an isolated nucleic acid molecule of the present invention includes a nucleic acid that is at least about 75%, preferably at least about 80%, more preferably at least about 85%, even more preferably at least about 90%, still more preferably at least about 95%, and even more preferably at least about 99%, identical to a chicken-derived lysozyme gene
20 expression control region-encoding nucleic acid molecule as depicted in SEQ ID NO: 67.

In another embodiment of the present invention, an avian lysozyme gene expression control region gene or nucleic acid molecule can be an allelic variant of SEQ ID NO: 67.

25 The present invention also contemplates the use of antisense nucleic acid molecules that are designed to be complementary to a coding strand of a nucleic acid (i.e., complementary to an mRNA sequence) or, alternatively, complimentary to a 5' or 3' untranslated region of the mRNA. Another use of synthetic nucleotides is as primers (DNA or RNA) for a polymerase chain reaction (PCR), ligase chain reaction
30 (LCR), or the like.

Synthesized nucleotides can be produced in variable lengths. The number of bases synthesized will depend upon a variety of factors, including the desired use for the probes or primers. Additionally, sense or anti-sense nucleic acids or oligonucleotides can be chemically synthesized using modified nucleotides to increase the biological stability of the molecule or of the binding complex formed between the anti-sense and sense nucleic acids. For example, acridine substituted nucleotides can be synthesized. Protocols for designing isolated nucleotides, nucleotide probes, and/or nucleotide primers are well-known to those of ordinary skill, and can be purchased commercially from a variety of sources (*e.g.*, Sigma Genosys, The Woodlands, TX or The Great American Gene Co., Ramona, CA).

The nucleic acid sequence of a chicken lysozyme gene expression control region nucleic acid molecule (SEQ ID NO: 67) of the present invention allows one skilled in the art to, for example, (a) make copies of those nucleic acid molecules by procedures such as, but not limited to, insertion into a cell for replication by the cell, by chemical synthesis or by procedures such as PCR or LCR, (b) obtain nucleic acid molecules which include at least a portion of such nucleic acid molecules, including full-length genes, full-length coding regions, regulatory control sequences, truncated coding regions and the like, (c) obtain lysozyme gene expression control region nucleic acid homologs in other avian species such as, but not limited to, turkey, duck, goose, quail, pheasant, parrot, finch, ratites including ostrich, emu and cassowary and, (d) to obtain isolated nucleic acids capable of hybridizing to an avian lysozyme gene expression control region nucleic acid and be used to detect the presence of nucleic acid-related sequences by complementation between the probe and the target nucleic acid.

Such nucleic acid homologs can be obtained in a variety of ways including by screening appropriate expression libraries with antibodies of the present invention, using traditional cloning techniques to screen appropriate libraries, amplifying appropriate libraries or DNA using oligonucleotide primers of the present invention in a polymerase chain reaction or other amplification method, and screening public and/or private databases containing genetic sequences using nucleic acid molecules of

the present invention to identify targets. Examples of preferred libraries to screen, or from which to amplify nucleic acid molecules, include but are not limited to mammalian BAC libraries, genomic DNA libraries, and cDNA libraries. Similarly, preferred sequence databases useful for screening to identify sequences in other
5 species homologous to chicken lysozyme gene expression control region include, but are not limited to, GenBank and the mammalian Gene Index database of The Institute of Genomics Research (TIGR).

Codon-optimized proteins

Another aspect of the present invention is a recombinant DNA molecule
10 comprising the novel isolated avian lysozyme gene expression control region of the present invention operably linked to a selected polypeptide-encoding nucleic acid insert, and which may express the nucleic acid insert when transfected to a suitable host cell, preferably an avian cell. The nucleic acid insert may be placed in frame with a signal peptide sequence, whereby translation initiation from the transcript may
15 start with the signal peptide and continue through the nucleic acid insert, thereby producing an expressed polypeptide having the desired amino acid sequence.

It is anticipated that the recombinant DNA, therefore, may further comprise a polyadenylation signal sequence that will allow the transcript directed by the novel lysozyme gene expression control region to proceed beyond the nucleic acid insert
20 encoding a polypeptide and allow the transcript to further comprise a 3' untranslated region and a polyadenylated tail. Any functional polyadenylation signal sequence may be linked to the 3' end of the nucleic acid insert including the SV40 polyadenylation signal sequence, bovine growth hormone adenylation sequence or the like, or derivatives thereof.

25 In one embodiment of the recombinant DNA of the present invention, the polyadenylation signal sequence is derived from the SV40 virus.

In another embodiment of the recombinant DNA of the present invention, the polyadenylation signal has the nucleic acid sequence SEQ ID NO: 68 or a variant thereof, as shown in Fig. 6.

It is further anticipated that the recombinant DNA of the present invention may further comprise the chicken lysozyme 3' domain SEQ ID NO: 69, or a variant thereof. The lysozyme 3' domain comprises a 3' untranslated region, a polyadenylation sequence and at least on 3' MAR.

5 Another aspect of the present invention is to provide nucleic acid sequences of a human interferon $\alpha 2b$ protein optimized for expression in avian cells, and derivatives and fragments thereof.

In derivatives of the human interferon $\alpha 2b$ protein of the present invention, for example, it is reasonable to expect that an isolated replacement of a leucine with an
10 isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar replacement of an amino acid with a structurally related amino acid (i.e. conservative mutations) will not have a major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino
15 acids can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion,
20 the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur-containing = cysteine and methionine. (see, for
25 example, "Biochemistry", 2nd ed, L. Stryer, ed., WH Freeman and Co.,1981). Peptides in which more than one replacement has taken place can readily be tested in the same manner.

One embodiment of the present invention is a recombinant DNA molecule comprising the isolated avian lysozyme gene expression control region of the present
30 invention, operably linked to a nucleic acid insert encoding a polypeptide, and a

polyadenylation signal sequence optionally operably linked thereto. It is contemplated that when the recombinant DNA is to be delivered to a recipient cell for expression therein, the sequence of the nucleic acid sequence may be modified so that the codons are optimized for the codon usage of the recipient species. For example, if the
5 recombinant DNA is transfected into a recipient chicken cell, the sequence of the expressed nucleic acid insert is optimized for chicken codon usage. This may be determined from the codon usage of at least one, and preferably more than one, protein expressed in a chicken cell. For example, the codon usage may be determined from the nucleic acid sequences encoding the proteins ovalbumin, lysozyme,
10 ovomucin and ovotransferrin of chicken.

In one embodiment of the recombinant DNA of the present invention, therefore, the nucleic acid insert encodes the human interferon $\alpha 2b$ polypeptide. Optimization of the sequence for codon usage elevates the level of translation in avian eggs. In this embodiment, the sequence (SEQ ID NO: 66) of the optimized human
15 interferon sequence is shown in Fig. 4.

In yet another embodiment of the present invention, the recombinant DNA comprises the isolated avian lysozyme gene expression control region operably linked to a nucleic acid encoding a human interferon $\alpha 2b$ and the SV40 polyadenylation sequence, the recombinant DNA having the nucleotide sequence SEQ ID NO: 65, as
20 shown in Fig. 3, or a variant thereof.

In still another embodiment of the present invention, the recombinant DNA comprises the isolated avian lysozyme gene expression control region operably linked to the nucleic acid encoding a polypeptide, and the chicken lysozyme 3' domain SEQ ID NO: 69. In one embodiment of the present invention, the nucleic acid insert is
25 SEQ ID NO: 66 encoding a human $\alpha 2b$ interferon, and the recombinant DNA construct has the nucleic acid sequence SEQ ID NO: 70.

The protein of the present invention may be produced in purified form by any known conventional techniques. For example, chicken cells may be homogenized and centrifuged. The supernatant is then subjected to sequential ammonium sulfate
30 precipitation and heat treatment. The fraction containing the protein of the present

invention is subjected to gel filtration in an appropriately sized dextran or polyacrylamide column to separate the proteins. If necessary, the protein fraction may be further purified by HPLC.

5 Recombinant nucleic acids, and expression thereof, under the control of an avian lysozyme promoter:

Another potentially useful application of the novel isolated lysozyme gene expression control region of the present invention is the possibility of increasing the amount of a heterologous protein present in a bird, (especially the chicken) by gene transfer. In most instances, a heterologous polypeptide-encoding nucleic acid insert
10 transferred into the recipient animal host will operably linked with the lysozyme gene expression control region, to allow the cell to initiate and continue production of the genetic product protein. A recombinant DNA molecule of the present invention can be transferred into the extra-chromosomal or genomic DNA of the host.

The recombinant DNA nucleic acid molecules of the present invention can be
15 delivered to cells using conventional recombinant DNA technology. The recombinant DNA molecule may be inserted into a cell to which the recombinant DNA molecule is heterologous (i.e. not normally present). Alternatively, as described more fully below, the recombinant DNA molecule may be introduced into cells which normally contain the recombinant DNA molecule, for example, to correct a deficiency in the expression
20 of a polypeptide, or where over-expression of the polypeptide is desired.

For expression in heterologous systems, the heterologous DNA molecule is inserted into the expression system or vector of the present invention in proper sense orientation and correct reading frame. The vector contains the necessary elements for the transcription and translation of the inserted protein-coding sequences, including
25 the novel isolated lysozyme gene expression control region.

U.S. Patent No. 4,237,224 to Cohen and Boyer, hereby incorporated by reference in its entirety, describes the production of expression systems in the form of recombinant plasmids using restriction enzyme cleavage and ligation with DNA ligase. These recombinant plasmids are then introduced to a cell by means of
30 transformation and replicated in cultures, including eukaryotic cells grown in tissue

culture.

One aspect of the present invention, therefore, is an expression vector suitable for delivery to a recipient cell for expression of the vector therein. It is contemplated to be within the scope of the present invention for the expression vector to comprise an isolated avian lysozyme gene expression control region operably linked to a nucleic acid insert encoding a polypeptide, and optionally a polyadenylation signal sequence. The expression vector of the present invention may further comprise a bacterial plasmid sequence, a viral nucleic acid sequence, or fragments or variants thereof that may allow for replication of the vector in a suitable host.

The novel isolated avian lysozyme gene expression control region of the present invention (SEQ ID NO: 67) and a polypeptide-encoding nucleic acid sequence operably linked thereto, such as, for example, SEQ ID NO: 66 or a derivative or truncated variant thereof, and optionally a polyadenylation signal sequence such as, for example, SEQ ID NO: 68 or the chicken lysozyme 3' domain may be introduced into viruses such as vaccinia virus. Methods for making a viral recombinant vector useful for expressing a protein under the control of the lysozyme promoter are analogous to the methods disclosed in U.S. Patent Nos. 4,603,112; 4,769,330; 5,174,993; 5,505,941; 5,338,683; 5,494,807; 4,722,848; Paoletti, E., 1996, *Proc. Natl. Acad. Sci.*, 93: 11349-11353; Moss, B., 1996, *Proc. Natl. Acad. Sci.* 93: 11341-11348; Roizman, 1996, *Proc. Natl. Acad. Sci.* 93: 11307-11302; Frolov et al., 1996, *Proc. Natl. Acad. Sci.* 93: 11371-11377; Grunhaus et al., 1993, *Seminars in Virology* 3: 237-252 and U.S. Patent Nos. 5,591,639; 5,589,466; and 5,580,859 relating to DNA expression vectors, *inter alia*; the contents of which are incorporated herein by reference in their entireties.

Recombinant viruses can also be generated by transfection of plasmids into cells infected with virus. Suitable vectors include, but are not limited to, viral vectors such as lambda vector system λ gt11, λ gt WES.tB, Charon 4, and plasmid vectors such as pBR322, pBR325, pACYC177, pACYC184, pUC8, pUC9, pUC18, pUC19, pLG339, pR290, pKC37, pKC101, SV 40, pBluescript II SK +/- or KS +/- (see "Stratagene Cloning Systems" Catalog (1993) from Stratagene, La Jolla, Calif., which

is hereby incorporated by reference), pQE, pIH821, pGEX, pET series (see Studier, F.W. et. al., 1990, *Use of T7 RNA Polymerase to Direct Expression of Cloned Genes* in "Gene Expression Technology," vol. 185, which is hereby incorporated by reference in its entirety) and any derivatives thereof. Recombinant molecules can be introduced into cells via transformation, particularly transduction, conjugation, mobilization, or electroporation. The DNA sequences are cloned into the vector using standard cloning procedures in the art, as described by Maniatis et al., 1982, *Molecular Cloning: A Laboratory Manual*, Cold Springs Laboratory, Cold Springs Harbor, N.Y., which is hereby incorporated by reference in its entirety.

A variety of host-vector systems may be utilized to express the protein-encoding sequence(s). Primarily, the vector system must be compatible with the host cell used. The use of eukaryotic recipient host cells permits partial or complete post-translational modification such as, but not only, glycosylation and/or the formation of the relevant inter- or intra-chain disulfide bonds. Host-vector systems include but are not limited to the following: bacteria transformed with bacteriophage DNA, plasmid DNA, or cosmid DNA; microorganisms such as yeast containing yeast vectors; vertebrate cell systems infected with virus (e.g., vaccinia virus, adenovirus, etc.); insect cell systems infected with virus (e.g., baculovirus) or avian embryonic cells inoculated with the recombinant nucleic acid. The expression elements of these vectors vary in their strength and specificities. Depending upon the host-vector system utilized, any one of a number of suitable transcription and translation elements can be used.

Once the novel isolated lysozyme gene expression control region of the present invention has been cloned into a vector system, it is ready to be incorporated into a host cell. Such incorporation can be carried out by the various forms of transformation noted above, depending upon the vector/host cell system. Suitable host cells include, but are not limited to, bacteria, virus, yeast, mammalian cells, and the like. Alternatively, it is contemplated that the incorporation of the DNA of the present invention into a recipient cell may be by any suitable method such as, but not limited to, viral transfer, electroporation, gene gun insertion, sperm mediated transfer

to an ovum, microinjection and the like.

Another aspect of the present invention, therefore, is a method of expressing a heterologous polypeptide in a eukaryotic cell by transfecting the cell with a recombinant DNA comprising an avian lysozyme gene expression control region
5 operably linked to a nucleic acid insert encoding a polypeptide and, optionally, a polyadenylation signal sequence, and culturing the transfected cell in a medium suitable for expression of the heterologous polypeptide under the control of the avian lysozyme gene expression control region.

In one embodiment of the method of the present invention, the recipient
10 eukaryotic cell is derived from an avian. In one embodiment, the avian is a chicken.

Yet another aspect of the present invention is a eukaryotic cell transformed with an expression vector according to the present invention and described above. In one embodiment of the present invention, the transformed cell is a chicken oviduct cell and the nucleic acid insert comprises the chicken lysozyme gene expression
15 control region, a nucleic acid insert encoding a human interferon $\alpha 2b$ and codon optimized for expression in an avian cell, and an SV40 polyadenylation sequence.

It is contemplated that the transfected cell according to the present invention may be transiently transfected, whereby the transfected recombinant DNA or expression vector may not be integrated into the genomic nucleic acid. It is further
20 contemplated that the transfected recombinant DNA or expression vector may be stably integrated into the genomic DNA of the recipient cell, thereby replicating with the cell so that each daughter cell receives a copy of the transfected nucleic acid. It is still further contemplated for the scope of the present invention to include a transgenic animal producing a heterologous protein expressed from a transfected nucleic acid
25 according to the present invention.

In one embodiment of the present invention, the transgenic animal is an avian selected from a turkey, duck, goose, quail, pheasant, ratite, an ornamental bird or a feral bird. In another embodiment, the avian is a chicken and the heterologous protein produced under the transcriptional control of the isolated avian lysozyme gene

expression control region according to the present invention is produced in the white of an egg.

Viral vector cell transformation:

An exemplary approach for the *in vivo* introduction of a nucleic acid encoding the subject novel isolated lysozyme gene expression control region into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA, encoding the gene product. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells that have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. Recombinant retrovirus can be constructed in the part of the retroviral coding sequence (gag, pol, env) that has been replaced by nucleic acid encoding a lysozyme gene expression control region, thereby rendering the retrovirus replication defective. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Ausubel et al., 1989, "Current Protocols in Molecular Biology," Sections 9.10-9.14, Greene Publishing Associates, and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM, all of which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include psiCrip, psiCre, psi2 and psiAm.

Furthermore, it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234, WO94/06920, and WO94/11524). For instance, strategies for the modification of the infection spectrum of retroviral vectors include coupling antibodies specific for cell surface antigens to the viral env protein (Roux et al., 1989, *Proc. Natl. Acad. Sci.* 86:

9079-9083; Julan et al., 1992, *J. Gen. Virol.* 73: 3251-3255 and Goud et al., 1983, *Virology* 163: 251-254) or coupling cell surface ligands to the viral env proteins (Neda et al., 1991, *J. Biol. Chem.* 266: 14143-14146)(all of which are incorporated herein by reference in their entirety). Coupling can be in the form of the chemical cross-
5 linking with a protein or other moiety (e.g., lactose to convert the env protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g., single-chain antibody/env fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an ecotropic vector into an amphotropic vector.

10 Another viral gene delivery system useful in the present invention utilizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes a gene product of interest, but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle (see, for example, Berkner et al., 1988, *BioTechniques* 6: 616; Rosenfeld et al., 1991, *Science* 252: 43 1434; and Rosenfeld et al., 1992, *Cell* 68: 143-155, all of which are incorporated herein by reference in their
15 entirety). Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. The virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect the
20 spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis in situations where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Most replication-defective adenoviral vectors
25 currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3 genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al., 1979, *Cell* 16:683; Berkner et al., *supra*; and Graham et al., 1991, pp. 109-127 in "Methods in Molecular Biology," vol. 7, E. J. Murray, ed., Humana, Clifton, N.J., all of which are incorporated herein by reference in their
30 entirety). Expression of an inserted gene such as, for example, encoding the human

interferon $\alpha 2b$, can be under control of the exogenously added lysozyme gene expression control region sequences.

Yet another viral vector system useful for delivery of, for example, the subject avian lysozyme gene expression control region operably linked to a nucleic acid encoding a polypeptide, is the adeno-associated virus (AAV). Vectors containing as little as 300 base pairs of AAV can be packaged and can integrate. Space for exogenous DNA is limited to about 4.5 kb. An AAV vector, such as that described in Tratschin et al., 1985, *Mol. Cell. Biol.* 5: 3251-3260, can be used to introduce DNA into cells. A variety of nucleic acids have been introduced into different cell types using AAV vectors (see, for example, Hermonat et al., 1984, *Proc. Natl. Acad. Sci.* 81: 6466-6470; Tratschin et al., 1985, *Mol. Cell. Biol.* 4: 2072-2081; Wondisford et al., 1988, *Mol. Endocrinol.* 2: 32-39; Tratschin et al., 1984, *J. Virol.* 51: 611-619; and Flotte et al., 1993, *J. Biol. Chem.* 268: 3781-3790, all of which are incorporated herein by reference in their entireties).

Non-viral expression vectors:

Most non-viral methods of gene transfer rely on normal mechanisms used by eukaryotic cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the subject lysozyme gene expression control region and operably linked polypeptide-encoding nucleic acid by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In a representative embodiment, a nucleic acid comprising the novel isolated lysozyme gene expression control region of the present invention can be entrapped in liposomes bearing positive charges on their surface (*e.g.*, lipofectins) and (optionally) which are tagged with antibodies against cell surface antigens of the target tissue (Mizuno et al., 1992, *NO Shinkei Geka* 20: 547-551; PCT publication WO91/06309; Japanese patent application 1047381; and European patent publication EP-A-43075, all of which are incorporated herein by reference in their entireties).

In similar fashion, the gene delivery system comprises an antibody or cell

surface ligand that is cross-linked with a gene binding agent such as polylysine (see, for example, PCT publications WO93/04701, WO92/22635, WO92/20316, WO92/19749, and WO92/06180, all of which are incorporated herein by reference in their entireties). It will also be appreciated that effective delivery of the subject
5 nucleic acid constructs via receptor-mediated endocytosis can be improved using agents which enhance escape of gene from the endosomal structures. For instance, whole adenovirus or fusogenic peptides of the influenza HA gene product can be used as part of the delivery system to induce efficient disruption of DNA-containing endosomes (Mulligan et al., 1993, *Science* 260-926; Wagner et al., 1992, *Proc. Natl.*
10 *Acad. Sci.* 89: 7934; and Christiano et al., 1993, *Proc. Natl. Acad. Sci.* 90: 2122, all of which are incorporated herein by reference in their entireties). It is further contemplated that a recombinant DNA molecule comprising the novel isolated lysozyme gene expression control region of the present invention may be delivered to a recipient host cell by other non-viral methods including by gene gun, microinjection,
15 sperm-mediated transfer, or the like.

Transgenic animals:

Another aspect of the present invention concerns transgenic animals, such as chickens, having a transgene comprising the novel isolated lysozyme gene expression control region of the present invention and which preferably (though optionally)
20 express a heterologous gene in one or more cells in the animal. Suitable methods for the generation of transgenic avians having heterologous DNA incorporated therein are described, for example, in WO 99/19472 to Ivarie et al.; WO 00/11151 to Ivarie et al.; and WO 00/56932 to Harvey et al., all of which are incorporated herein by reference in their entirety.

25 In various embodiments of the present invention, the expression of the transgene may be restricted to specific subsets of cells, tissues or developmental stages utilizing, for example, *cis*-acting sequences acting on the lysozyme gene expression control region of the present invention and which control gene expression in the desired pattern. Tissue-specific regulatory sequences and conditional regulatory
30 sequences can be used to control expression of the transgene in certain spatial

patterns. Moreover, temporal patterns of expression can be provided by, for example, conditional recombination systems or prokaryotic transcriptional regulatory sequences. The inclusion of a 5' MAR region in the novel isolated lysozyme gene expression control region of the present invention may allow the heterologous expression unit to escape the chromosomal positional effect (CPE) and therefore be expressed at a more uniform level in transgenic tissues that received the transgene by a route other than through germ line cells.

One embodiment of the present invention, therefore, is a transgenic avian having a heterologous polynucleotide sequence comprising a nucleic acid insert encoding the heterologous polypeptide and operably linked to the novel isolated avian lysozyme gene expression control region, the lysozyme gene expression control region comprising at least one 5' matrix attachment region, an intrinsically curved DNA region, at least one transcription enhancer, a negative regulatory element, at least one hormone responsive element, at least one avian CR1 repeat element, and a proximal lysozyme promoter and signal peptide-encoding region.

In an embodiment of the present invention, the transgenic avian is selected from a chicken, a turkey, a duck, a goose, a quail, a pheasant, a ratite, an ornamental bird or a feral bird.

In another embodiment of the present invention, the transgenic avian is a chicken.

In still another embodiment of the transgenic avian of the present invention, the transgenic avian includes an avian lysozyme gene expression control region comprising the nucleic acid sequence in SEQ ID NO: 67, or a degenerate variant thereof.

In yet another embodiment of the transgenic avian of the present invention, the transgenic avian further comprises a polyadenylation signal sequence.

In still yet another embodiment of the transgenic avian of the present invention, the polyadenylation signal sequence is derived from the SV40 virus.

In an embodiment of the transgenic avian of the present invention, the polyadenylation signal sequence comprises the nucleic acid sequence in SEQ ID NO: 68, or a degenerate variant thereof.

5 In still another embodiment of transgenic avian of the present invention, the transgenic avian comprises the chicken lysozyme 3' domain having the nucleic acid sequence SEQ ID NO: 69.

In another embodiment of the transgenic avian of the present invention, the nucleic acid insert encoding a polypeptide has a codon complement optimized for protein expression in an avian.

10 In yet another embodiment of the transgenic avian of the present invention, the nucleic acid insert encodes an interferon $\alpha 2b$ polypeptide.

In still another embodiment of the transgenic avian of the present invention, the nucleic acid insert encoding an interferon $\alpha 2b$ polypeptide comprises the sequence in SEQ ID NO: 66, or a degenerate variant thereof.

15 In one embodiment of the transgenic avian of the present invention, the transgenic avian comprises the nucleotide sequence in SEQ ID NO: 65, or a degenerate variant thereof.

In yet another embodiment of the transgenic avian of the present invention, the transgenic avian comprises the nucleotide sequence in SEQ ID NO: 70, or a
20 degenerate variant thereof.

In another embodiment of the transgenic avian of the present invention, the transgenic avian produces the heterologous polypeptide in the serum or an egg white.

In another embodiment of the transgenic avian of the present invention, the transgenic avian produces the heterologous polypeptide in an egg white.

25 The present invention is further illustrated by the following examples, which are provided by way of illustration and should not be construed as limiting. The contents of all references, published patents and patents cited throughout the present application are hereby incorporated by reference in their entireties.

Example 1: Construction of lysozyme promoter plasmids

The chicken lysozyme gene expression control region was isolated by PCR amplification. Ligation and reamplification of the fragments thereby obtained yielded a contiguous nucleic acid construct comprising the chicken lysozyme gene expression control region operably linked to a nucleic acid sequence optimized for codon usage in the chicken (SEQ ID NO: 66) and encoding a human interferon $\alpha 2b$ polypeptide optimized for expression in an avian cell.

White Leghorn Chicken (*Gallus gallus*) genomic DNA was PCR amplified using the primers 5pLMAR2 (SEQ ID NO: 1) (see Fig. 1) and LE-6.1kbrev1 (SEQ ID NO: 2) in a first reaction, and Lys-6.1 (SEQ ID NO: 3) and LysE1rev (SEQ ID NO: 4) as primers in a second reaction. PCR cycling steps were: denaturation at 94°C for 1 minute; annealing at 60°C for 1 minute; extension at 72°C for 6 minutes, for 30 cycles using TAQ PLUS PRECISION™ DNA polymerase (Stratagene, LaJolla, CA). The PCR products from these two reactions were gel purified, and then united in a third PCR reaction using only 5pLMAR2 (SEQ ID NO: 1) and LysE1rev (SEQ ID NO: 4) as primers and a 10-minute extension period. The resulting DNA product was phosphorylated, gel-purified, and cloned into the *EcoR* V restriction site of the vector pBluescript KS, resulting in the plasmid p12.0-lys.

p12.0-lys was used as a template in a PCR reaction with primers 5pLMAR2 (SEQ ID NO: 1) and LYSBSU (SEQ ID NO: 5) and a 10 minute extension time. The resulting DNA was phosphorylated, gel-purified, and cloned into the *EcoR* V restriction site of pBluescript KS, forming plasmid p12.0lys-B.

p12.0lys-B was restriction digested with *Not* I and *Bsu*36 I, gel-purified, and cloned into *Not* I and *Bsu*36 I digested pCMV-LysSPIFNMM, resulting in p12.0-lys-LSPIFNMM. p12.0-lys-LSPIFNMM was digested with *Sal* I and the *Sal*Ito*Not*I primer (SEQ ID NO: 6) was annealed to the digested plasmid, followed by *Not* I digestion. The resulting 12.5 kb *Not* I fragment, comprising the lysozyme promoter region linked to IFNMAGMAX-encoding region and an SV40 polyadenylation signal sequence, was gel-purified and ligated to *Not* I cleaved and dephosphorylated pBluescript KS, thereby forming the plasmid pAVIJCR-A115.93.1.2. The lysozyme

promoter/IFN construct contained in plasmid pAVIJCR-A115.93.1.2 was sequenced as described in Example 3.

Example 2: Construction of plasmids which contain the 3' lysozyme domain.

The plasmid pAVIJCR-A115.93.1.2 was restriction digested with *FseI* and blunt-ended with T4 DNA polymerase. The linearized, blunt-ended pAVIJCR-A115.93.1.2 plasmid was then digested with *XhoI* restriction enzyme, followed by treatment with alkaline phosphatase. The resulting 15.4 kb DNA band containing the lysozyme 5'matrix attachment region (MAR) and -12.0 kb lysozyme promoter driving expression of a human interferon was gel purified by electroelution.

The plasmid pIIIlys was restriction digested with *MluI*, then blunt-ended with the Klenow fragment of DNA polymerase. The linearized, blunt-ended pIIIlys plasmid was digested with *XhoI* restriction enzyme and the resulting 6 kb band containing the 3' lysozyme domain from exon 3 to the 3' end of the 3' MAR was gel purified by electroelution. The 15.4 kb band from pAVIJCR-A115.93.1.2 and the 6 kb band from pIIIlys were ligated with T4 DNA ligase and transformed into STBL4 cells (Invitrogen Life Technologies, Carlsbad, CA) by electroporation. The resulting 21.3 kb plasmids from two different bacterial colonies were named pAVIJCR-A212.89.2.1 and pAVIJCR-A212.89.2.3 respectively.

Example 3: Sequencing Reactions

Plasmid DNA (pAVIJCR-A115.93.1.2) produced as described in Example 1 was purified with QIAGEN™ columns (Qiagen, Valencia, CA). Sequencing reactions were performed according to the Applied Biosystems (Foster City, CA) protocol for BIGDYE™ Terminators, version 2.0, using an ABI 373 Stretch sequencer. Sequencing primers used are listed in Fig. 1, and a schematic diagram illustrating the sequencing reactions using the different primers is shown in Fig. 2. Sequence data was analyzed with SEQUENCHER™ software, version 4.0 (Gene Codes Corp., Ann Arbor, MI).

Example 4: Complete lysozyme promoter and IFNMAGMAX sequences

The complete nucleotide sequence (SEQ ID NO: 65), shown in Fig. 3, of the 12.5 kb chicken lysozyme promoter region/IFNMAGMAX construct spans the 5'

matrix attachment region (5' MAR), through the lysozyme signal peptide, to the sequence encoding the gene IFNMAGMAX and the subsequent polyadenylation signal sequence. The IFNMAGMAX nucleic acid sequence (SEQ ID NO: 66), shown in Fig. 4, encoded human interferon $\alpha 2b$ (IFN) that had been synthesized based on a codon usage table compiled from the four most abundantly expressed hen egg white proteins ovalbumen, ovotransferrin, ovomucoid and lysozyme. The expressed IFN $\alpha 2b$ sequence within plasmid pAVIJCRA115.93.1.2 functioned as a reporter gene for lysozyme promoter activity. This plasmid construct may also be used for production of interferon $\alpha 2b$ in the egg white of transgenic chickens. The isolated sequence of the 11.94 kb chicken lysozyme promoter region (SEQ ID NO: 67) alone is shown in Fig. 5. The sequence of the SV40 polyadenylation signal sequence (SEQ ID NO: 68) is shown in Fig. 6.

Example 5: Basic Local Alignment Search Tool (BLAST) Analysis of the Complete Lysozyme Promoter Sequence (SEQ ID NO: 65)

The complete 12.5 kb lysozyme promoter/IFNMAGMAX sequence (SEQ ID NO: 65) was submitted to the National Center for Biotechnology Information for BLAST alignments with database sequences. Percent identities between the lysozyme promoter sequence (SEQ ID NO: 67, included within SEQ ID NO: 65) and corresponding known lysozyme promoter features are shown in Table II below:

Table II. BLAST Results of the Complete 12.0 kb Lysozyme Promoter Sequence

Description of DNA element	Coordinates in this sequence	GenBank accession number	% identity
5' matrix attachment region	1-237, 261-1564	AJ277960	96
5' matrix attachment region	1-237, 261-1564	X98408	96
5' matrix attachment region	1564-1912 1930-2015	X84223	99
Intrinsically curved DNA	2011-2671	X52989	98
Transcription enhancer (-6.1 kb)	5848-5934	<u>Grewal et al.</u> , 1992	100

Transcription enhancer (E-2.7 kb)	9160-9329	X05461	98
Negative regulatory element	9325-9626	X05463	98
Hormone response element	9621-9666 9680-10060	X12509	99
CR1 chicken repeat element	10576-10821, 10926-11193	U88211, K02907	87
Transcription enhancer (E-0.2 kb)	11655-11797	X05462	100
Proximal promoter and lysozyme signal peptide	11563-11877	M12532	100
Proximal promoter and lysozyme signal peptide	11424-11938	J00886	99

Features that have been previously identified as individual elements isolated from other component elements of the lysozyme promoter region include the 5' MAR, three transcription enhancers, a hormone-responsive element, and a chicken repeat 1 (CR1) element. The IFNMAGMAX sequence (SEQ ID NO: 66) extended from nucleotide positions 11946 to 12443 of SEQ ID NO: 65, shown in Fig. 3.

Example 6: Expression in Transfected Cultured Avian Oviduct Cells of Human Interferon α 2b Regulated by the 12kb Lysozyme Promoter

The oviduct was removed from a Japanese quail (*Coturnix coturnix japonica*) and the magnum portion minced and enzymatically dissociated with 0.8 mg/ml collagenase (Sigma Chemical Co., St. Louis, MO) and 1.0 mg/ml dispase (Roche Molecular Biochemicals, Indianapolis, IN) by shaking and titrating for 30 minutes at 37°C. The cell suspension was then filtered through sterile surgical gauze, washed three times with F-12 medium (Life Technologies, Grand Island, NY) by centrifugation at 200 x g, and resuspended in OPTIMEM™ (Life Technologies) such that the OD₆₀₀ was approximately 2. Cell suspension (300 μ l) was plated per well of a 24-well dish. For each transfection, 2.5 μ l of DMRIE-C liposomes (Life

Technologies) and 1 µg of DNA were preincubated for 15 minutes at room temperature in 100 µl of OPTIMEM™, and then added to the oviduct cells. Cells with DNA/liposomes were incubated for 5 hours at 37°C in 5% CO₂. Next, 0.75 ml of DMEM (Life Technologies) supplemented with 15% fetal bovine serum (FBS) (Atlanta Biologicals, Atlanta, GA), 2X penicillin/streptomycin (Life Technologies), 10⁻⁶ M insulin (Sigma), 10⁻⁸ M β-estradiol (Sigma), and 10⁻⁷ M corticosterone (Sigma) was added to each well, and incubation was continued for 72 hours. Medium was then harvested and centrifuged at 110 x g for 5 minutes. The supernatant was analyzed by ELISA for human interferon α2b content.

The human interferon α2b contents of medium derived from cultured oviduct cells transfected with either the -12.0 kb IFN plasmid (pAVIJCR-A115.93.1.2) or the negative control plasmid pCMV-EGFP as shown in Fig. 7. Bars to the right of the figure represent the standards for the IFN ELISA.

Example 7: Transfection of chicken HD11 cells with pAVIJCR-A212.89.2.1 and pAVIJCR-A212.89.2.3

Chicken cells transfected with plasmids having the 3 lysozyme domain linked to a nucleic acid expressing human α2b interferon express the heterologous polypeptide. Chicken myelomonocytic HD11 cells were transfected with plasmid pAVIJCR-A212.89.2.1 and pAVIJCR-A212.89.2.3 to test the functionality of the plasmids. One million HD11 cells were plated per each well of a 24-well dish. The next day, HD11 cells were transfected with 1 µg of plasmid DNA per 4 µl of LipofectAMINE 2000 (Invitrogen Life Technologies). For comparison, independent wells were also transfected with the parent vector pAVIJCR-A115.93.1.2. After 5 hours of transfection, the cell medium was changed with fresh medium. 48 hours later, cell medium was harvested by centrifugation at 110 X g for 5 min and assayed for human interferon by ELISA (PBL Biomedicals, Flanders, NJ).

The transfected cells expressed the heterologous human α2b interferon at least to the level seen with a plasmid not having the 3 lysozyme domain operably linked to the human α2b interferon encoding nucleic acid, as shown in Fig. 10.

Example 8: Expression of human α2b interferon in a transgenic avian platform

The plasmid pAVIJCR-A115.93.1.2 (containing the -12.0 kb lysozyme promoter controlling expression of human interferon α -2b) was purified with a Qiagen Plasmid Maxi Kit (Qiagen, Valencia, CA), and 100 micrograms of the plasmid were restriction digested with *NotI* restriction enzyme. The digested DNA was
 5 phenol/ CHCl_3 extracted and ethanol precipitated. Recovered DNA was resuspended in 1mM Tris-HCl (pH 8.0) and 0.1mM EDTA, then placed overnight at 4° C. DNA was quantified by spectrophotometry and diluted to the appropriate concentration. DNA samples were resuspended in 0.25 M KCl and bound with a SV40 T antigen nuclear localization signal peptide (NLS peptide, amino acid sequence
 10 CGGPKKKRKVG-NH₂; SEQ ID NO.: 71) by adding the NLS at a peptide:DNA molar ratio of 100:1 (Collas and Alestrom, 1996, Mol. Reprod. Develop. 45: 431-438; the contents of which is incorporated by reference in its entirety).

Cytoplasmic Microinjection of DNA: Approximately two nanoliters of DNA were injected into the germinal disk of stage I White Leghorn embryos obtained two hours
 15 after oviposition of the previous egg. DNA amounts per injection ranged from 10 picograms to 400 picograms. Injected embryos were surgically transferred to recipient hens via ovum transfer according to the method of Christmann et al (PCT/US01/26723 to Christmann et al; the contents of which is hereby incorporated by reference in its entirety)

20 *Analysis of chick blood DNA by PCR for IFN transgene:* Whole blood from one week old chicks was collected with heparinized capillary tubes. Red blood cell (RBC) nuclei were released and washed with lysis buffer solution. DNA's from RBC nuclei were extracted by digestion with proteinase K (1mg/ml) and precipitated with ethanol. Purified DNA was resuspended in 1mM Tris-HCl (pH 8.0) and 0.1mM EDTA and
 25 quantitated. Genomic DNA samples were analysed by PCR using primers LYS051 (SEQ ID NO.: 72) for 5'-TGCATCCTTCAGCACTTGAG-3' and IFN-3 (SEQ ID NO.: 73) for 5'-AACTCCTCTTGAGGAAAGCC-3'. This primer set amplifies a 584 bp region of the transgene carried by the pAVIJCR-A115.93.1.2 plasmid. Three hundred nanograms of genomic DNA were added to a 50 μ l reaction mixture (1 X
 30 Promega PCR Buffer with 1.5mM MgCl₂, 200 μ M of each dNTP, 5 μ M primers) and

1.25 units of Taq DNA polymerase (Promega). The reaction mixtures were heated for 4 minutes at 94°C, and then amplified for 34 cycles at 94°C for 1 min, 60°C for 1 min and 72°C for 1 min. The samples were heated in a final cycle for 4 minutes at 72°C. PCR products were detected on a 0.8% agarose gel with ethidium bromide staining.

- 5 *Human interferon α -2b expression in chick blood by ELISA:* One week after hatch, blood was collected from chicks using heparinized capillary tubes; added to an equal volume of phosphate buffered saline and centrifuged at 200 x g. 100 microliters of the supernatant were assayed by human IFN ELISA (PBL Biomedical Laboratories, New Brunswick, New Jersey).
- 10 *Human interferon α -2b expression in egg white of transgenic hens:* once hens reached sexual maturity and began to lay (approximately 22 – 24 weeks of age), eggs were collected and egg white assayed by ELISA using human IFN ELISA (PBL Biomedical Laboratories, New Brunswick, New Jersey) according to manufacturer's instructions.
- 15 *Results of PCR and ELISA analysis of blood and egg white:* Table III below summarizes results of PCR and ELISA analysis.

Table III: Analysis of Transgene presence and Interferon Expression

Bird #	Method	PCR (Blood)	ELISA (Blood)	ELISA (egg white)	# Birds Tested
8305	- NLS	+	+	NA (male)	69 (2.5%)
8340	- NLS	-	-	+	
AA123	+ NLS	+	+	NA (immature)	43 (9%)
AA61	+ NLS	+	+	"	
AA105	+ NLS	-	+	"	
AA115	+ NLS	+	-	"	

-NLS: DNA injected without NLS peptide; + NLS: DNA injected with NLS peptide; NA: not applicable

- 20 As shown in Table III, one bird (#8305) of 69 produced using microinjection of DNA without the NLS peptide, was positive for both presence of the transgene and expression of interferon in the blood. Because this bird is a male, he will be bred to a non-transgenic hen to examine germline transmission of the transgene. Figure 11
- 25 demonstrates expression of human interferon in the blood of #8305, as compared to standards. Figure 12 illustrates PCR results from the serum of for several birds,

including #8305, obtained at different intervals after hatch. As can be seen in lanes 4, 5, 11, and 12, positive signal was seen indicating the presence of the transgene at two different collection periods.

Other positives were seen in birds produced under microinjection of DNA covalently linked to the NLS peptide as described above. Table III illustrates 4 birds (AA123, AA61, AA105 and AA115) out of 43 tested that were PCR positive, ELISA positive or both. Expression levels of human IFN in AA61, as compared to standards, is also illustrated in Figure 11. Males will be bred to determine germline transmission, and eggs collected from transgenic females to assay for IFN expression, as described above, as chicks reach sexual maturity.

Example 9: Expression of a human monoclonal antibody (Mab) in a transgenic avian platform

Transgenic chickens were produced as described in Example 8 above, except that two distinct constructs were coinjected into Stage 1 embryos. The constructs comprised the 12 kb lysozyme promoter, as described above in Example 4, driving either a heavy chain or light chain of a human monoclonal antibody against CTLA-4 (WO 01/14424 A2 to *Korman et al.*; the contents of which is incorporated herein in its entirety). ELISA analysis of serum, conducted as described above in Example 8, is summarized below:

Table IV. ELISA analysis of Mab expression in hatched birds:

Bird #	Method	ELISA (serum)	ELISA (egg white)	# Birds Tested
214	+ NLS	+	NA (immature)	13
228	+ NLS	+	"	

+ NLS: DNA injected with NLS peptide; NA: not available

Results indicate that two birds of the thirteen tested to date, #214 and #228, are positive for Mab expression in the serum.

Although preferred embodiments of the invention have been described using specific terms, devices, and methods, such description is for illustrative purposes only. The

words used are words of description rather than of limitation. It is to be understood that changes and variations may be made by those of ordinary skill in the art without departing from the spirit or the scope of the present invention, which is set forth in the following claims. In addition, it should be understood that aspects of the various

5 embodiments may be interchanged both in whole or in part.

What is Claimed Is:

1. An isolated nucleic acid comprising an isolated avian lysozyme gene expression control region comprising:
 - (a) at least one 5' matrix attachment region;
 - 5 (b) an intrinsically curved DNA region;
 - (c) at least one transcription enhancer;
 - (d) a *negative regulatory element*;
 - (e) at least one hormone responsive element;
 - (f) at least one avian CR1 repeat element; and
 - 10 (g) a proximal lysozyme promoter and signal peptide-encoding region.
2. The isolated nucleic acid of Claim 1, wherein the avian is selected from the group consisting of a chicken, a turkey, a duck, a goose, a quail, a pheasant, a ratite, an ornamental bird or a feral bird.
3. The isolated nucleic acid of Claim 1, wherein the lysozyme gene expression control region comprises the nucleic acid sequence in SEQ ID NO: 67, or a degenerate variant thereof.
- 15 4. The isolated nucleic acid of Claim 1 comprising a sequence at least 75% identical to SEQ ID NO: 67.
5. The isolated nucleic acid of Claim 1 comprising a sequence at least 95% identical to SEQ ID NO: 67.
- 20 6. The isolated nucleic acid of Claim 1 comprising a sequence at least 99% identical to SEQ ID NO: 67.
7. A recombinant DNA molecule comprising an isolated avian lysozyme gene expression control region operably linked to a nucleic acid insert encoding a polypeptide, wherein the lysozyme gene expression control region comprises:
 - 25 (a) at least one 5' matrix attachment region;
 - (b) an intrinsically curved DNA region;
 - (c) at least one transcription enhancer;
 - (d) a *negative regulatory element*;
 - 30 (e) *at least one hormone responsive element*;

- (f) at least one avian CR1 repeat element; and
- (g) a proximal lysozyme promoter and signal peptide-encoding region.
8. The recombinant DNA molecule of Claim 7, wherein the avian is selected from the group consisting of a chicken, a turkey, a duck, a goose, a quail, a pheasant, a ratite, an ornamental bird or a feral bird.
9. The recombinant DNA molecule of Claim 7, wherein the lysozyme gene expression control region comprises the nucleic acid sequence in SEQ ID NO: 67, or a degenerate variant thereof.
10. The recombinant DNA molecule of Claim 7, further comprising a polyadenylation signal sequence, said polyadenylation signal sequence optionally comprising the nucleic acid sequence in SEQ ID NO: 68, or a degenerate variant thereof.
11. The recombinant DNA molecule of Claim 7, wherein the nucleic acid insert encodes an interferon $\alpha 2b$ polypeptide; said polypeptide optionally comprising the sequence in SEQ ID NO: 66, or a degenerate variant thereof.
12. The recombinant DNA molecule of Claim 7 having the nucleotide sequence in SEQ ID NO: 65, or a degenerate variant thereof.
13. The recombinant DNA molecule of Claim 7 having the nucleotide sequence in SEQ ID NO: 70, or a degenerate variant thereof.
14. An expression vector that integrates into a host cell and comprising an isolated avian lysozyme gene expression control region operably linked to a nucleic acid insert encoding a polypeptide, wherein the expression control region directs production of a transcript, wherein the lysozyme gene expression control region comprises:
- (a) at least one 5' matrix attachment region;
- (b) an intrinsically curved DNA region;
- (c) at least one transcription enhancer;
- (d) a negative regulatory element;
- (e) at least one hormone responsive element;
- (f) at least one avian CR1 repeat element; and

- (g) a proximal lysozyme promoter and signal peptide-encoding region.
15. The expression vector of Claim 14, wherein the lysozyme gene expression control region comprises the nucleic acid sequence in SEQ ID NO: 67, or a degenerate variant thereof.
- 5 16. The expression vector of Claim 14 wherein the nucleic acid insert encodes an interferon $\alpha 2b$ polypeptide; said polypeptide optionally comprising the sequence in SEQ ID NO: 66, or a degenerate variant thereof.
17. The expression vector of Claim 16, wherein the nucleic acid insert encoding an interferon $\alpha 2b$ polypeptide comprises the sequence in SEQ ID NO: 66, or a degenerate variant thereof.
- 10 18. The expression vector of Claim 14 having the nucleotide sequence in SEQ ID NO: 65, or a degenerate variant thereof.
19. The expression vector of Claim 14 comprising the nucleotide sequence in SEQ ID NO: 70, or a degenerate variant thereof.
- 15 20. A method of expressing a heterologous polypeptide in a host cell, comprising the steps of:
- (a) transfecting a eukaryotic cell with a recombinant DNA molecule as claimed in Claim 5, thereby generating a transfected cell;
- (b) culturing the transfected cell in a medium suitable for expression of
- 20 a heterologous polypeptide under the control of an avian lysozyme gene expression control region encoded by the recombinant DNA molecule.
21. A eukaryotic cell transformed with the expression vector according to Claim 14, or a progeny of the cell, wherein the cell or the progeny thereof expresses a heterologous polypeptide.
- 25 22. The eukaryotic cell of Claim 21, wherein the cell is an oviduct cell of a chicken or a cultured cell.
23. The eukaryotic cell of Claim 21, wherein the expression vector has a nucleic acid insert encoding a polypeptide selected from the group consisting of an

- interferon or a monoclonal antibody, and wherein said nucleic acid insert has a codon complement optimized for protein expression in an avian.
24. The eukaryotic cell of Claim 23, wherein the nucleic acid insert encoding the interferon $\alpha 2b$ polypeptide comprises the sequence in SEQ ID NO: 66, or a degenerate variant thereof.
25. The eukaryotic cell of Claim 23, wherein the insert encoding the interferon $\alpha 2b$ polypeptide comprises the sequence in SEQ ID NO: 70, or a degenerate variant thereof.
26. A transgenic avian having a heterologous polynucleotide sequence comprising a nucleic acid insert encoding the heterologous polypeptide and operably linked to an avian lysozyme gene expression control region, wherein the lysozyme gene expression control region comprises:
- (a) at least one 5' matrix attachment region;
 - (b) an intrinsically curved DNA region;
 - (c) at least one transcription enhancer;
 - (d) a negative regulatory element;
 - (e) at least one hormone responsive element;
 - (f) at least one avian CR1 repeat element; and
 - (g) a proximal lysozyme promoter and signal peptide-encoding region.
27. The transgenic avian of Claim 26, wherein the avian is selected from the group consisting of a chicken, a turkey, a duck, a goose, a quail, a pheasant, a ratite, an ornamental bird or a feral bird.
28. The transgenic avian of Claim 26, wherein the lysozyme gene expression control region comprises the nucleic acid sequence in SEQ ID NO: 67, or a degenerate variant thereof.
29. The transgenic avian of Claim 26, wherein the nucleic acid insert encodes an protein or a monoclonal antibody.
30. The transgenic avian of Claim 26, wherein the nucleic acid insert encoding the an interferon $\alpha 2b$ polypeptide comprises the sequence in SEQ ID NO: 66, or a degenerate variant thereof.

31. The transgenic avian of Claim 26 having the nucleotide sequence in SEQ ID NO: 65, or a degenerate variant thereof.
32. The transgenic avian of Claim 26 having the nucleotide sequence in SEQ ID NO: 70, or a degenerate variant thereof.
- 5 33. The transgenic avian of Claim 26 wherein the transgenic avian produces the heterologous polypeptide in the serum or an egg white.

SEQ ID NOS: 1-64

5pLMAR2 TGCCGCCTTCTTTGATATTC	SEQ ID NO: 1
LE-6.1kbrev1 TTGGTGGTAAGGCCTTTTTG	SEQ ID NO: 2
Lys-6.1 CTGGCAAGCTGTCAAAAACA	SEQ ID NO: 3
LysE1rev CAGCTCACATCGTCCAAAGA	SEQ ID NO: 4
LYSBSU CCCCCCCCTAAGGCAGCCAGGGGCAGGAAGCAAA	SEQ ID NO: 5
SalItoNotI TCGAGCGGCCGC	SEQ ID NO: 6
T7 TAATACGACTCACTATAGGG	SEQ ID NO: 7
lys61enfor1 CGTGGTGATCAAATCTTTGTG	SEQ ID NO: 8
lys61enrev1 AGGAGGGCACAGTAGGGATC	SEQ ID NO: 9
5MARfor1 GTGGCCTGTGTCTGTGCTT	SEQ ID NO: 10
IFN-3rev AACTCCTCTTGAGGAAAGCC	SEQ ID NO: 11
lys001rev TCCTGTTTGGGATGAATGGT	SEQ ID NO: 12
lys002for CTCTCAGAATGCCCAACTCC	SEQ ID NO: 13
lys003for TGTATTGGTCTCCCTCCTGC	SEQ ID NO: 14
lys005for TGTTGAAATTGCAGTGTGGC	SEQ ID NO: 15
lys006rev TGACAATGCAAATTTGGCTC	SEQ ID NO: 16

lys007for GATATCCTTG CAGTGCCCAT	SEQ ID NO: 17
lys008rev GGACAAGCAAGTGCATCAGA	SEQ ID NO: 18
lys009for CTGATGTGCTTCAGCTCTGC	SEQ ID NO: 19
lys010rev TCCATGGTGGTCAAACAGAA	SEQ ID NO: 20
lys011for GTACTAGACCAGGCAGCCCA	SEQ ID NO: 21
lys012rev GTGGGAAGTACCACATTGGC	SEQ ID NO: 22
lys013for CGCTCAGGAGAAAGTGAACC	SEQ ID NO: 23
lys014rev CGGTTTTGCCTTTGTGTTTT	SEQ ID NO: 24
lys015rev AAATGCTCGATTTCATTGGG	SEQ ID NO: 25
lys016rev GCCAATCAGACTGCATTTCA	SEQ ID NO: 26
lys017rev AACCGCTGAATGGAACAGTC	SEQ ID NO: 27
lys018for ACACGCACATATTTTGCTGG	SEQ ID NO: 28
lys019rev CAGGAGCTGGATTCCCTTCAG	SEQ ID NO: 29
lys020for AAAGGATGCAGTCCCAAATG	SEQ ID NO: 30
lys021rev GCCCCTAGACTCCATCTTCC	SEQ ID NO: 31
lys022rev ATTTGCTGTGGTGGATGTGA	SEQ ID NO: 32
lys024for CCTTGCAGTCCTTGTTTGT	SEQ ID NO: 33

lys025rev ATGATCCTTCTGATGGGCTG	SEQ ID NO: 34
lys026rev ACAGTGATAGCACAAGGGGG	SEQ ID NO: 35
lys027rev GTAAACAGCTGCAACAGGCA	SEQ ID NO: 36
lys028rev CAACACAAAAGTTGGACAGCA	SEQ ID NO: 37
lys029rev TTTGCAGATGAGACGTTTGC	SEQ ID NO: 38
lys030rev CCACAAGTTCTTGTGGGGC	SEQ ID NO: 39
lys031rev ATCAATCCATGCCAGTAGCC	SEQ ID NO: 40
lys032rev GTTTAAGGCCCTTCCAATC	SEQ ID NO: 41
lys033for GAGAGGGGGTTGGGTGTATT	SEQ ID NO: 42
lys034for ACAGTGGAAGCATTCAGGG	SEQ ID NO: 43
lys037for CCAATGCCTTTGGTTCTGAT	SEQ ID NO: 44
lys038for AAAACACAAAGGCAAAACCG	SEQ ID NO: 45
lys039rev CTAAGCCTCGCCAGTTTCAA	SEQ ID NO: 46
lys040rev TGCCATGAAAACCCTACTGA	SEQ ID NO: 47
lys041for GGAATGTACCCTCAGCTCCA	SEQ ID NO: 48
lys042rev CCTCTTTAGGAGGCCAGCTT	SEQ ID NO: 49
lys043rev AAGATGATCAGAGGGCTGGA	SEQ ID NO: 50

lys044rev GCAGCGCTGGTAATCTTCAT	SEQ ID NO: 51
lys045for CTTCAGATCCCAGGAAGTGC	SEQ ID NO: 52
lys046for TTCCTGCCTTACATTCTGGG	SEQ ID NO: 53
lys047for CCCCTGCAGGCTTAGAAAG	SEQ ID NO: 54
lys048for AGTTCTCCATAGCGGCTGAA	SEQ ID NO: 55
lys051for TGCATCCTTCAGCACTTGAG	SEQ ID NO: 56
lys052rev GCAGGAGGGAGACCAATACA	SEQ ID NO: 57
lys053for TGCACAAGGATGTCTGGGTA	SEQ ID NO: 58
lys054for TCCTAGCAACTGCGGATTTT	SEQ ID NO: 59
lys056for TCTTCCATGTTGGTGACAGC	SEQ ID NO: 60
lys058for CCCCCTTGTGCTATCACTGT	SEQ ID NO: 61
lys059for CTGACAGACATCCCAGCTCA	SEQ ID NO: 62
lys060for AAGTTGTGCTTCTGCGTGTG	SEQ ID NO: 63
lys061for TTGTTCTGCTGTTCTCCTCT	SEQ ID NO: 64

Fig. 1

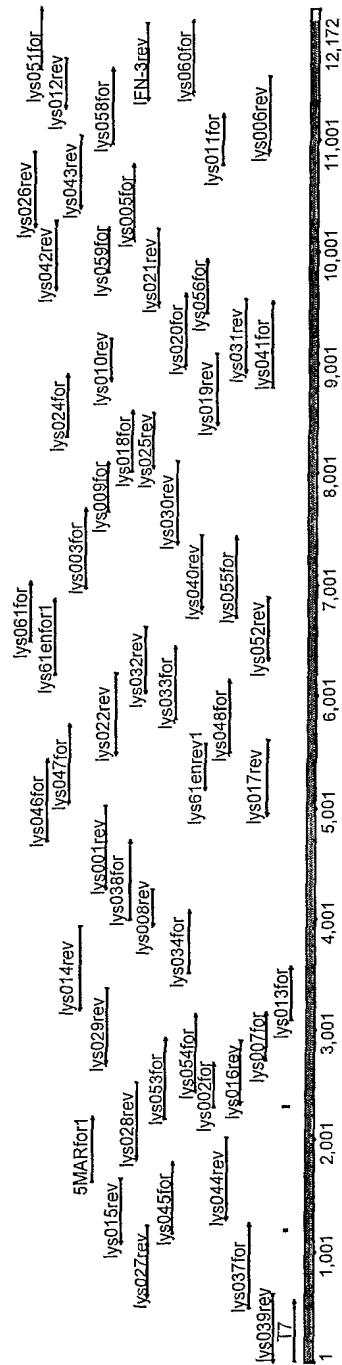


Fig.2

TGCCGCCTTC	TTTGATATTC	ACTCTGTTGT	ATTTTCATCTC	TTCTTGCCGA	TGAAAGGATA	60
TAACAGTCTG	TATAACAGTC	TGTGAGGAAA	TACTTGGTAT	TTCTTCTGAT	CAGTGTTTTT	120
ATAAGTAATG	TTGAATATTG	GATAAGGCTG	TGTGTCCTTT	GTCTTGGGAG	ACAAAGCCCCA	180
CAGCAGGTGG	TGGTTGGGGT	GGTGGCAGCT	CAGTGACAGG	AGAGGTTTTT	TTGCCTGTTT	240
TTTTTTTTTTT	TTTTTTTTTTT	AAGTAAGGTG	TTCTTTTTTTC	TTAGTAAATT	TTCTACTGGA	300
CTGTATGTTT	TGACAGGTCA	GAAACATTTT	TTCAAAAGAA	GAACCTTTTTG	GAAACTGTAC	360
AGCCCTTTTTC	TTTCATTCCC	TTTTTGCTTT	CTGTGCCAAT	GCCTTTGGTT	CTGATTGCAT	420
TATGGAAAAC	GTTGATCGGA	ACTTGAGGTT	TTTATTTATA	GTGTGGCTTG	AAAGCTTGGA	480
TAGCTGTTGT	TACACGAGAT	ACCTTATTAA	GTTTAGGCCA	GCTTGATGCT	TTATTTTTTTC	540
CCTTTGAAGT	AGTGAGCGTT	CTCTGGTTTT	TTTCCTTTGA	AACTGGTGAG	GCTTAGATTT	600
TTCTAATGGG	ATTTTTTTACC	TGATGATCTA	GTTGCATACC	CAAATGCCTG	TAAATGTTTT	660
CCTAGTTAAC	ATGTTGATAA	CTTCGGATTT	ACATGTTGTA	TATACTTGTC	ATCTGTGTTT	720
CTAGTAAAAA	TATATGGCAT	TTATAGAAAT	ACGTAATTCC	TGATTTCCCT	TTTTTTTTATC	780
TCTATGCTCT	GTGTGTACAG	GTCAAACAGA	CTTCACTCCT	ATTTTTATTT	ATAGAATTTT	840
ATATGCAGTC	TGTCGTTGGT	TCTTGTGTTG	TAAGGATACA	GCCTTAAATT	TCCTAGAGCG	900
ATGCTCAGTA	AGGCGGGTTG	TCACATGGGT	TCAAATGTAA	AACGGGCACG	TTTGGCTGCT	960
GCCTTCCCGA	GATCCAGGAC	ACTAAACTGC	TTCTGCACTG	AGGTATAAAT	CGCTTCAGAT	1020
CCCAGGGAAG	TGCAGATCCA	CGTGCAATAT	CTTAAAGAAG	AATGAATACT	TTCTAAAATA	1080
TTTTGGCATA	GGAAGCAAGC	TGCATGGATT	TGTTTTGGGAC	TTAAATTATT	TTGGTAACGG	1140
AGTGCATAGG	TTTTAAACAC	AGTTGCAGCA	TGCTAACGAG	TCACAGCGTT	TATGCAGAAG	1200
TGATGCCTGG	ATGCCTGTTG	CAGCTGTTTA	CGGCACTGCC	TTGCAGTGAG	CATTGCAGAT	1260
AGGGGTGGGG	TGCTTTGTGT	CGTGTTCCCA	CACGCTGCCA	CACAGCCACC	TCCCGGAACA	1320
CATCTCACCT	GCTGGGTACT	TTTCAAACCA	TCTTAGCAGT	AGTAGATGAG	TTACTATGAA	1380
ACAGAGAAGT	TCCTCAGTTG	GATATTCTCA	TGGGATGTCT	TTTTTCCCAT	GTTGGGCAAA	1440
GTATGATAAA	GCATCTCTAT	TTGTAAATTA	TGCACTTGTT	AGTTCCTGAA	TCCTTTCTAT	1500
AGCACCACCT	ATTGCAGCAG	GTGTAGGCTC	TGGTGTGGCC	TGTGTCTGTG	CTTCAATCTT	1560
TTAAAGCTTC	TTTGGAATAA	CACTGACTTG	ATTGAAGTCT	CTTGAAGATA	GTAAACAGTA	1620
CTTACCTTTG	ATCCCAATGA	AATCGAGCAT	TTCAAGTTGTA	AAAGAATTCC	GCCTATTCAT	1680
ACCATGTAAAT	GTAATTTTAC	ACCCCCAGTG	CTGACACTTT	GGAATATATT	CAAGTAATAG	1740
ACTTTGGCCT	CACCTCTTTG	TGTACTGTAT	TTTGTAATAG	AAAATATTTT	AAACTGTGCA	1800
TATGATTATT	ACATTATGAA	AGAGACATTC	TGCTGATCTT	CAAATGTAAG	AAAATGAGGA	1860
GTGCGTGTGC	TTTTATAAAT	ACAAGTGATT	GCAAATTAGT	GCAGGTGTCC	TTAAAAAAA	1920
AAAAAAAAAAG	TAATATAAAA	AGGACCAGGT	GTTTTACAAG	TGAAATACAT	TCCTATTTGG	1980
TAAACAGTTA	CATTTTTTATG	AAGATTACCA	GCGCTGCTGA	CTTTCTAAAC	ATAAGGCTGT	2040
ATTGTCTTCC	TGTACCATTG	CATTTCTCTA	TTCCCAATTT	GCACAAGGAT	GTCTGGGTAA	2100
ACTATTCAAG	AAATGGCTTT	GAAATACAGC	ATGGGAGCTT	GTCTGAGTTG	GAATGCAGAG	2160
TTGCACTGCA	AAATGTCAGG	AAATGGATGT	CTCTCAGAAT	GCCCAACTCC	AAAGGATTTT	2220
ATATGTGTAT	ATAGTAAGCA	GTTTCCTGAT	TCCAGCAGGC	CAAAGAGTCT	GCTGAATGTT	2280
GTGTTGCCGG	AGACCTGTAT	TTCTCAACAA	GGTAAGATGG	TATCCTAGCA	ACTGCGGATT	2340
TTAATACATT	TTCAGCAGAA	GTAAGTATG	AATCTCTACC	TTTAGGGATC	GTTTCATCAT	2400
TTTTAGATGT	TATACTTGAA	ATACTGCATA	ACTTTTAGCT	TTTATGGGTT	CCTTTTTTTC	2460
AGCCTTTAGG	AGACTGTTAA	GCAATTTGCT	GTCCAACTTT	TGTGTTGGTC	TTAAACTGCA	2520
ATAGTAGTTT	ACCTTGATTT	GAAGAAATAA	AGACCATTTT	TATATTAAAA	AATACTTTTG	2580
TCTGTCTTCA	TTTTGACTTG	TCTGATATCC	TTGCAAGTCC	CATTATGTCA	GTTCTGTCAG	2640
ATATTCAGAC	ATCÀAAACTT	AACGTGAGCT	CAGTGGAGTT	ACAGCTGCGG	TTTTGATGCT	2700
GTTATTATTT	CTGAAACTAG	AAATGATGTT	GTCTTCATCT	GCTCATCAAA	CACTTCATGC	2760
AGAGTGTAAG	GCTAGTGAGA	AATGCATACA	TTTATTGATA	CTTTTTTAAA	GTCAACTTTT	2820
TATCAGATTT	TTTTTTTCATT	TGGAAATATA	TTGTTTTCTA	GACTGCATAG	CTTCTGAATC	2880
TGAAATGCAG	TCTGATTGGC	ATGAAGAAGC	ACAGCACTCT	TCATCTTACT	TAAACTTCAT	2940
TTTGGAATGA	AGGAAGTTAA	GCAAGGGCAC	AGGTCCATGA	AATAGAGACA	GTGCGCTCAG	3000
GAGAAAGTGA	ACCTGGATTT	CTTTGGCTAG	TGTTCTAAAT	CTGTAGTGAG	GAAAGTAACA	3060
CCCGATTCCCT	TGAAAGGGCT	CCAGCTTTAA	TGCTTCCAAA	TTGAAGGTGG	CAGGCAACTT	3120
GGCCACTGGT	TATTTACTGC	ATTATGTCTC	AGTTTCGCAG	CTAACCTGGC	TTCTCCACTA	3180
TTGAGCATGG	ACTATAGCCT	GGCTTCAGAG	GCCAGGTGAA	GGTTGGGATG	GGTGGAAGGA	3240
GTGCTGGGCT	GTGGCTGGGG	GGACTGTGGG	GACTCCAAGC	TGAGCTTGGG	GTGGGCAGCA	3300

CAGGGAAAAG	TGTGGGTAAC	TATTTTTTAAG	TACTGTGTTG	CAAACGTCTC	ATCTGCAAAT	3360
ACGTAGGGTG	TGTACTCTCG	AAGATTAACA	GTGTGGGTTC	AGTAATATAT	GGATGAATTC	3420
ACAGTGGAAG	CATTCAAGGG	TAGATCATCT	AACGACACCA	GATCATCAAG	CTATGATTGG	3480
AAGCGGTATC	AGAAGAGCGA	GGAAAGGTAAG	CAGTCTTCAT	ATGTTTTCCC	TCCACGTAAA	3540
GCAGTCTGGG	AAAGTAGCAC	CCCTTGAGCA	GAGACAAGGA	AATAATTCAG	GAGCATGTGC	3600
TAGGAGAACT	TTCTTGCTGA	ATTCTACTTG	CAAGAGCTTT	GATGCCTGGC	TTCTGGTGCC	3660
TTCTGCAGCA	CCTGCAAGGC	CCAGAGCCTG	TGGTGAGCTG	GAGGGAAAGA	TTCTGCTCAA	3720
GTCCAAGCTT	CAGCAGGTCA	TTGTCTTTGC	TTCTTCCCCC	AGCACTGTGC	AGCAGAGTGG	3780
AACTGATGTC	GAAGCCTCCT	GTCCACTACC	TGTTGCTGCA	GGCAGACTGC	TCTCAGAAAA	3840
AGAGAGCTAA	CTCTATGCCA	TAGTCTGAAG	GTAAAATGGG	TTTTAAAAAA	GAAAAACACAA	3900
AGGCAAAACC	GGCTGCCCCA	TGAGAAGAAA	GCAGTGGTAA	ACATGGTAGA	AAAGGTGCAG	3960
AAGCCCCCAG	GCAGTGTGAC	AGGCCCTCC	TGCCACCTAG	AGGCGGGAAC	AAGCTTCCCT	4020
GCCTAGGGCT	CTGCCCCGCA	AGTGCGTGTT	TCTTTGGTGG	GTTTTGTTTG	GCCTTGGTT	4080
TTGAGATTTA	GACACAAGGG	AAGCCTGAAA	GGAGGTGTTG	GGCACTATTT	TGGTTGTAA	4140
AGCCTGTACT	TCAAATATAT	ATTTTGTGAG	GGAGGTGAGC	GAATTGGCCA	ATTTAAAAATA	4200
AAGTTGCAAG	AGATTGAAGG	CTGAGTAGTT	GAGAGGGTAA	CACGTTTAAT	GAGATCTTCT	4260
GAAACTACTG	CTTCTAAACA	CTTGTTTGAG	TGGTGAGACC	TTGGATAGGT	GAGTGCTCTT	4320
GTTACATGTC	TGATGCACTT	GCTTGTCTCT	TTCCATCCAC	ATCCATGCAT	TCCACATCCA	4380
CGCATTTGTC	ACTTATCCCA	TATCTGTCAT	ATCTGACATA	CCTGTCTCTT	CGTCACTTGG	4440
TCAGAAGAAA	CAGATGTGAT	AATCCCCAGC	CGCCCCAAGT	TTGAGAAGAT	GGCAGTTGCT	4500
TCTTTCCCTT	TTTCCTGCTA	AGTAAGGATT	TTCTCCTGGC	TTTGACACCT	CACGAAATAG	4560
TCTTCCTGCC	TTACATTCTG	GGCATTATTT	CAAATATCTT	TGGAGTGCGC	TGCTCTCAAG	4620
TTTGTGTCTT	CCTACTCTTA	GAGTGAATGC	TCTTAGAGTG	AAAGAGAAGG	AAGAGAAGAT	4680
GTTGGCCGCA	GTTCTCTGAT	GAACACACCT	CTGAATAATG	GCCAAAGGTG	GGTGGGTTTC	4740
TCTGAGGAAC	GGGCAGCGTT	TGCCCTTGAA	AGCAAGGAGC	TCTGCGGAGT	TGCAGTTATT	4800
TTGCAACTGA	TGGTGGAAC	GGTGCTTAAA	GCAGATTCCC	TAGGTTCCCT	GCTACTTCTT	4860
TTCCCTCTTG	GCAGTCAGTT	TATTTCTGAC	AGACAAACAG	CCACCCCCAC	TGCAGGCTTA	4920
GAAAGTATGT	GGCTCTGCCT	GGGTGTGTTA	CAGCTCTGCC	CTGGTGAAAG	GGGATAAAAA	4980
CGGGCACCAT	TCATCCCAAA	CAGGATCCTC	ATTCATGGAT	CAAGCTGTAA	GGAACCTGGG	5040
CTCCAACCTC	AAAACATTAA	TTGGAGTACG	AATGTAATTA	AAACTGCATT	CTCGCATTCC	5100
TAAGTCATTT	AGTCTGGACT	CTGCAGCATG	TAGGTCGGCA	GCTCCCACTT	TCTCAAAGAC	5160
CACTGATGGA	GGAGTAGTAA	AAATGGAGAC	CGATTCAAGAA	CAACCAACGG	AGTGTTGCCG	5220
AAGAAACTGA	TGGAAATAAT	GCATGAATTG	TGTGGTGGAC	ATTTTTTTTA	AATACATAAA	5280
CTACTTCAAA	TGAGGTCGGA	GAAGGTCAGT	GTTTTATTAG	CAGCCATAAA	ACCAGGTGAG	5340
CGAGTACCAT	TTTTCTCTAC	AAGAAAAACG	ATTCTGAGCT	CTGCGTAAGT	ATAAGTTCTC	5400
CATAGCGGCT	GAAGCTCCCC	CCTGGCTGCC	TGCCATCTCA	GCTGGAGTGC	AGTGCCATTT	5460
CCTTGGGGTT	TCTCTCACAG	CAGTAATGGG	ACAATACTTC	ACAAAAATTC	TTTCTTTTCC	5520
TGTCATGTGG	GATCCCTACT	GTGCCCTCCT	GGTTTTACGT	TACCCCTGA	CTGTTCCATT	5580
CAGCGGTTTG	GAAAGAGAAA	AAGAATTTGG	AAATAAAACA	TGTCTACGTT	ATCACCTCCT	5640
CCAGCATTTT	GGTTTTTAAT	TATGTCAATA	ACTGGCTTAG	ATTTGGAAAT	GAGAGGGGGT	5700
TGGGTGTATT	ACCGAGGAAC	AAAGGAAGGC	TTATATAAAC	TCAAGTCTTT	TATTTAGAGA	5760
ACTGGCAAGC	TGTCAAAAAC	AAAAAGGCCT	TACCACCAAA	TTAAGTGAAT	AGCCGCTATA	5820
GCCAGCAGGG	CCAGCACGAG	GGATGGTGCA	CTGCTGGCAC	TATGCCACGG	CCTGCTTGTG	5880
ACTCTGAGAG	CAACTGCTTT	GGAAATGACA	GCACTTGGTG	CAATTTCCCT	TGTTTCAGAA	5940
TGCGTAGAGC	GTGTGCTTGG	CGACAGTTTT	TCTAGTTAGG	CCACTTCTTT	TTTCTTCTC	6000
TCCTCATTCT	CCTAAGCATG	TCTCCATGCT	GGTAATCCCA	GTCAAGTGAA	CGTTCAAACA	6060
ATGAATCCAT	CACTGTAGGA	TTCTCGTGGT	GATCAAATCT	TTGTGTGAGG	TCTATAAAAT	6120
ATGGAAGCTT	ATTTATTTTT	CGTTCTTCCA	TATCAGTCTT	CTCTATGACA	ATTCACATCC	6180
ACCACAGCAA	ATTAAAGGTG	AAGGAGGCTG	GTGGGATGAA	GAGGGTCTTC	TAGCTTTACG	6240
TTCTTCCTTG	CAAGGCCACA	GGAAATGCT	GAGAGCTGTA	GAATACAGCC	TGGGGTAAGA	6300
AGTTCAGTCT	CCTGCTGGGA	CAGCTAACCG	CATCTTATAA	CCCCTCTGA	GACTCATCTT	6360
AGGACCAAAT	AGGGTCTATC	TGGGGTTTTT	GTTCCCTGCTG	TTCCCTCTGG	AAGGCTATCT	6420
CACATATTTA	CTGCTCCCAC	GGTTACAAAC	CAAAGATACA	GCCTGAATTT	TTTCTAGGCC	6480
ACATTACATA	AATTTGACCT	GGTACCAATA	TTGTTCTCTA	TATAGTTATT	TCCTTCCCCA	6540
CTGTGTTTAA	CCCCTTAAGG	CATTCAGAAC	AACTAGAATC	ATAGAATGGT	TTGGATTGGA	6600

AGGGGCGCTTA	AACATCATCC	ATTTCCAACC	CTCTGCCATG	GGCTGCTTGC	CACCCACTGG	6660
CTCAGGCTGC	CCAGGGCCCC	ATCCAGCCTG	GCCTTGAGCA	CCTCCAGGGA	TGGGGCACCC	6720
ACAGCTTCTC	TGGGCAGCCT	GTGCCAACAC	CTCACCCTC	TCTGGGTAAA	GAATTCTCTT	6780
TTAACATCTA	ATCTAAATCT	CTTCTCTTTT	AGTTTAAAGC	CATTCTCTTT	TTTCCCGTTG	6840
CTATCTGTCC	AAGAAATGTG	TATTGGTCTC	CCTCCTGCTT	ATAAGCAGGA	AGTACTGGAA	6900
GGCTGCAGTG	AGGTCTCCCC	ACAGCCTTCT	CTTCTCCAGG	CTGAACAAGC	CCAGCTCCTT	6960
CAGCCTGTCT	TCGTAGGAGA	TCATCTTAGT	GGCCCTCCTC	TGGACCCATT	CCAACAGTTC	7020
CACGGCTTTC	TTGTGGAGCC	CCAGGTCTGG	ATGCAGTACT	TCAGATGGGG	CCTTACAAAG	7080
GCAGAGCAGA	TGGGGACAAT	CGCTTACCCC	TCCCTGCTGG	CTGCCCCTGT	TTTGATGCAG	7140
CCCAGGGTAC	TGTTGGCCTT	TCAGGCTCCC	AGACCCCTTG	CTGATTTGTG	TCAAGCTTTT	7200
CATCCACCAG	AACCCACGCT	TCCTGGTTAA	TACTTCTGCC	CTCACTTCTG	TAAGCTTGTT	7260
TCAGGAGACT	TCCATTCTTT	AGGACAGACT	GTGTTACACC	TACCTGCCCT	ATTCTTGCAT	7320
ATATACATTT	CAGTTCATGT	TTCCTGTAAC	AGGACAGAAT	ATGTATTCTT	CTAACAAAAA	7380
TACATGCAGA	ATTCCTAGTG	CCATCTCAGT	AGGGTTTTCA	TGGCAGTATT	AGCACATAGT	7440
CAATTTGCTG	CAAGTACCTT	CCAAGCTGCG	GCCTCCCAT	AATCCTGTAT	TTGGGATCAG	7500
TTACCTTTTG	GGGTAAGCTT	TTGTATCTGC	AGAGACCCTG	GGGGTTCTGA	TGTGCTTCAG	7560
CTCTGCTCTG	TTCTGACTGC	ACCATTTTCT	AGATCACCCA	GTTGTTCTCT	TACAACCTTC	7620
TTGTCCTCCA	TCCTTTCCCA	GCTTGTATCT	TTGACAAATA	CAGGCCTATT	TTTGTGTTTG	7680
CTTCAGCAGC	CATTTAATTC	TTCAGTGTCA	TCTTGTTCTG	TTGATGCCAC	TGGAACAGGA	7740
TTTTTCAGCAG	TCTTGCAAAG	AACATCTAGC	TGAAAACTTT	CTGCCATTCA	ATATTCTTAC	7800
CAGTTCCTTCT	TGTTTGAGGT	GAGCCATAAA	TTACTAGAAC	TTCGTCACTG	ACAAGTTTAT	7860
GCATTTTATT	ACTTCTATTA	TGTACTTACT	TTGACATAAC	ACAGACACGC	ACATATTTTG	7920
CTGGGATTTT	CACAGTGTCT	CTGTGTCCTT	CACATGGTTT	TACTGTCATA	CTTCCGTTAT	7980
AACCTTGCCA	ATCTGCCCAG	CTGCCCATCA	CAAGAAAAGA	GATTCTTTTT	TTATTACTTC	8040
TCTTCAGCCA	ATAAACAAAA	TGTGAGAAGC	CCAAACAAGA	ACTTGTGGGG	CAGGCTGCCA	8100
TCAAGGGAGA	GACAGCTGAA	GGGTTGTGTA	GCTCAATAGA	ATTAAGAAAT	AATAAAGCTG	8160
TGTTCAGACAG	TTTTCCTGGA	TTTATACAGG	CACGCCCCAA	GCCAGAGAGG	CTGTCTGCCA	8220
AGGCCACCTT	GCAGTCCTTG	GTTTGTAAGA	TAGTCAATAG	GTAACTTTTT	TGGTGAATTG	8280
CGTGGAGAAT	CATGATGGCA	GTTCTTGCTG	TTTACTATGG	TAAGATGCTA	AAATAGGAGA	8340
CAGCAAAGTA	ACACTTGCTG	CTGTAGGTGC	TCTGCTATCC	AGACAGCGAT	GGCACTCGCA	8400
CACCAAGATG	AGGGATGCTC	CCAGCTGACG	GATGCTGGGG	CAGTAACAGT	GGGTCCCATG	8460
CTGCCTGCTC	ATTAGCATCA	CCTCAGCCCT	CACCAGCCCA	TCAGAAGGAT	CATCCCAAGC	8520
TGAGGAAAGT	TGCTCATCTT	CTTCACATCA	TCAAACCTTT	GGCCTGACTG	ATGCCTCCCG	8580
GATGCTTAAA	TGTGGTCACT	GACATCTTTA	TTTTTCTATG	ATTTCAAGTC	AGAACCTCCG	8640
GATCAGGAGG	GAACACATAG	TGGGAATGTA	CCCTCAGCTC	CAAGGCCAGA	TCTTCCTTCA	8700
ATGATCATGC	ATGCTACTTA	GGAAGGTGTG	TGTGTGTGAA	TGTAGAATTG	CCTTTGTTAT	8760
TTTTTCTTCC	TGCTGTTCAGG	AACATTTTGA	ATACCAGAGA	AAAAGAAAAG	TGCTCTTCTT	8820
GGCATGGGAG	GAGTTGTAC	ACTTGCAAAA	TAAAGGATGC	AGTCCCAAAT	GTTTATAATC	8880
TCAGGGTCTG	AAGGAGGATC	AGAACTGTG	TATACAATTT	CAGGCTTCTC	TGAATGCAGC	8940
TTTTTGAAAGC	TGTTCTTGCC	CGAGGCAGTA	CTAGTCAGAA	CCCTCGGAAA	CAGGAACAAA	9000
TGTCTTCAAG	GTGCAGCAGG	AGGAAACACC	TTGCCCATCA	TGAAAGTGAA	TAACCACTGC	9060
CGCTGAAGGA	ATCCAGCTCC	TGTTTGAGCA	GGTGCTGCAC	ACTCCACAC	TGAAACAACA	9120
GTTTCATTTTT	ATAGGACTTC	CAGGAAGGAT	CTTCTTCTTA	AGCTTCTTAA	TTATGGTACA	9180
TCTCCAGTTG	GCAGATGACT	ATGACTACTG	ACAGGAGAAT	GAGGAACTAG	CTGGGAATAT	9240
TTCTGTTTGA	CCACCATGGA	GTCACCCATT	TCTTTACTGG	TATTTGGAAA	TAATAATTCT	9300
GAATTGCAAA	GCAGGAGTTA	GCGAAGATCT	TCATTTCTTC	CATGTTGGTG	ACAGCACAGT	9360
TCTGGCTATG	AAAGTCTGCT	TACAAGGAAG	AGGATAAAAA	TCATAGGGAT	AATAAATCTA	9420
AGTTTGAAGA	CAATGAGGTT	TTAGCTGCAT	TTGACATGAA	GAAATTGAGA	CCTCTACTGG	9480
ATAGCTATGG	TATTTACGTG	TCTTTTGTCT	TAGTTACTTA	TTGACCCAG	CTGAGGTCAA	9540
GTATGAACTC	AGGTCTCTCG	GGCTACTGGC	ATGGATTGAT	TACATACAAC	TGTAATTTTA	9600
GCAGTGATTT	AGGGTTTATG	AGTACTTTTG	CAGTAAATCA	TAGGGTTAGT	AATGTTAATC	9660
TCAGGGAAAA	AAAAAAAAG	CCAACCCTGA	CAGACATCCC	AGCTCAGGTG	GAAATCAAGG	9720
ATCACAGCTC	AGTGCGGTCC	CAGAGAACAC	AGGGACTCTT	CTTTTAGGAC	CTTTATGTAC	9780
AGGGCCTCAA	GATAACTGAT	GTTAGTCAGA	AGACTTTCCA	TTCTGGCCAC	AGTTCAGCTG	9840
AGGCAATCCT	GGAATTTTCT	CTCCGCTGCA	CAGTTCCAGT	CATCCCAGTT	TGTACAGTTC	9900

TGGCACTTTT	TGGGTCAGGC	CGTGATCCAA	GGAGCAGAAG	TTCCAGCTAT	GGTCAGGGAG	9960
TGCCTGACCG	TCCCAACTCA	CTGCACTCAA	ACAAAGGCGA	AACCACAAGA	GTGGCTTTTG	10020
TTGAAATTGC	AGTGTGGCCC	AGAGGGGCTG	CACCAGTACT	GGATTGACCA	CGAGGCAACA	10080
TTAATCCTCA	GCAAGTGCAA	TTTGCAGCCA	TTAAATTGAA	CTAACTGATA	CTACAATGCA	10140
ATCAGTATCA	ACAAGTGGTT	TGGCTTGGAA	GATGGAGTCT	AGGGGCTCTA	CAGGAGTAGC	10200
TACTCTCTAA	TGGAGTTGCA	TTTTGAAGCA	GGACACTGTG	AAAAGCTGGC	CTCCTAAAGA	10260
GGCTGCTAAA	CATTAGGGTC	AATTTTCCAG	TGCACCTTCT	GAAGTGTCTG	CAGTTCCCCA	10320
TGCAAAGCTG	CCCAAACATA	GCACTTCCAA	TTGAATACAA	TTATATGCAG	GCGTACTGCT	10380
TCTTGCCAGC	ACTGTCCTTC	TCAAATGAAC	TCAACAAACA	ATTTCAAAGT	CTAGTAGAAA	10440
GTAACAAGCT	TTGAATGTCA	TTAAAAAGTA	TATCTGCTTT	CAGTAGTTCA	GCTTATTTAT	10500
GCCCACTAGA	AACATCTTGT	ACAAGCTGAA	CACTGGGGCT	CCAGATTAGT	GGTAAAACCT	10560
ACTTTATACA	ATCATAGAAT	CATAGAATGG	CCTGGGTTGG	AAGGGACCCC	AAGGATCATG	10620
AAGATCCAAC	ACCCCCGCCA	CAGGCAGGGC	CACCAACCTC	CAGATCTGGT	ACTAGACCAG	10680
GCAGCCCAGG	GCTCCATCCA	ACCTGGCCAT	GAACACCTCC	AGGGATGGAG	CATCCACAAC	10740
CTCTCTGGGC	AGCCTGTGCC	AGCACCTCAC	CACCCTCTCT	GTGAAGAAGT	TTTCCCTGAC	10800
ATCCAATCTA	AGCCTTCCCT	CCTTGAGGTT	AGATCCACTC	CCCCTTGTGC	TATCACTGTC	10860
TACTCTTGTA	AAAAGTTGAT	TCTCCTCCTT	TTTGGAAGGT	TGCAATGAGG	TCTCCTTGCA	10920
GCCTTCTTCT	CTTCTGCAGG	ATGAACAAGC	CCAGCTCCCT	CAGCCTGTCT	TTATAGGAGA	10980
GGTGCTCCAG	CCCTCTGATC	ATCTTTGTGG	CCCTCCTCTG	GACCCGCTCC	AAGAGCTCCA	11040
CATCTTTCCT	GTACTGGGGG	CCCCAGGCCT	GAATGCAGTA	CTCCAGATGG	GGCCTCAAAA	11100
GAGCAGAGTA	AAGAGGGACA	ATCACCTTCC	TCACCCTGCT	GGCCAGCCCT	CTTCTGATGG	11160
AGCCCTGGAT	ACAACCTGGCT	TTCTGAGCTG	CAACTTCTCC	TTATCAGTTC	CACTATTAAA	11220
ACAGGAACAA	TACAACAGGT	GCTGATGGCC	AGTGCAGAGT	TTTTCACACT	TCTTCATTTT	11280
GGTAGATCTT	AGATGAGGAA	CGTTGAAGTT	GTGCTTCTGC	GTGTGCTTCT	TCCTCCTCAA	11340
ATACTCCTGC	CTGATACCTC	ACCCACCTG	CCACTGAATG	GCTCCATGGC	CCCCTGCAGC	11400
CAGGGCCCTG	ATGAACCCGG	CACTGCTTCA	GATGCTGTTT	AATAGCACAG	TATGACCAAG	11460
TTGCACCTAT	GAATACACAA	ACAATGTGTT	GCATCCTTCA	GCACTTGAGA	AGAAGAGCCA	11520
AATTTGCATT	GTCAGGAAAT	GGTTTAGTAA	TTCTGCCAAT	TAAAACTTGT	TTATCTACCA	11580
TGGCTGTTTT	TATGGCTGTT	AGTAGTGGTA	CACTGATGAT	GAACAATGGC	TATGCAGTAA	11640
AATCAAGACT	GTAGATATTG	CAACAGACTA	TAAAAATTCCT	CTGTGGCTTA	GCCAATGTGG	11700
TACTTCCCAC	ATTGTATAAG	AAATTTGGCA	AGTTTAGAGC	AATGTTTGAA	GTGTTGGGAA	11760
ATTTCTGTAT	ACTCAAGAGG	GCGTTTTTGA	CAACTGTAGA	ACAGAGGAAT	CAAAAGGGGG	11820
TGGGAGGAAG	TTAAAAGAAG	AGGCAGGTGC	AAGAGAGCTT	GCAGTCCCGC	TGTGTGTACG	11880
ACACTGGCAA	CATGAGGTCT	TTGCTAATCT	TGGTGCTTTG	CTTCTTGCCC	CTGGCTGCCT	11940
TAGGGTGCGA	TCTGCCTCAG	ACCCACAGCC	TGGGCAGCAG	GAGGACCCTG	ATGCTGCTGG	12000
CTCAGATGAG	GAGAATCAGC	CTGTTTAGCT	GCCTGAAGGA	TAGGCACGAT	TTTGGCTTTT	12060
CTCAAGAGGA	GTTTGGCAAC	CAGTTTCAGA	AGGCTGAGAC	CATCCCTGTG	CTGCACGAGA	12120
TGATCCAGCA	GATCTTTAAC	CTGTTTAGCA	CCAAGGATAG	CAGCGCTGCT	TGGGATGAGA	12180
CCCTGCTGGA	TAAGTTTAC	ACCGAGCTGT	ACCAGCAGCT	GAACGATCTG	GAGGCTTGCG	12240
TGATCCAGGG	CGTGGGCGTG	ACCGAGACCC	CTCTGATGAA	GGAGGATAGC	ATCCTGGCTG	12300
TGAGGAAGTA	CTTTCAGAGG	ATCACCTGT	ACCTGAAGGA	GAAGAAGTAC	AGCCCCTGCG	12360
CTTGGGAAGT	CGTGAGGGCT	GAGATCATGA	GGAGCTTTAG	CCTGAGCACC	AACCTGCAAG	12420
AGAGCTTGAG	GTCTAAGGAG	TAAAAAGTCT	AGAGTCGGGG	CGGCCGGCCG	CTTCGAGCAG	12480
ACATGATAAG	ATACATTGAT	GAGTTTGGAC	AAACCACAAC	TAGAATGCAG	TGAAAAAAT	12540
GCTTTATTTG	TGAAATTTGT	GATGCTATTG	CTTTATTTGT	AACCATTATA	AGCTGCAATA	12600
AACAAGTTAA	CAACAACAAT	TGCATTCAAT	TTATGTTTCA	GGTTCAGGGG	GAGGTGTGGG	12660
AGGTTTTTTA	AAGCAAGTAA	AACCTCTACA	AATGTGGTAA	AATCGATAAG	GATCCGTCGA	12720
GCGGCCCG	12728					

FIG. 3

SEQ ID NO: 66

TGCGATCTGC	CTCAGACCCA	CAGCCTGGGC	AGCAGGAGGA	CCCTGATGCT	GCTGGCTCAG	60
ATGAGGAGAA	TCAGCCTGTT	TAGCTGCCTG	AAGGATAGGC	ACGATTTTGG	CTTTCCTCAA	120
GAGGAGTTTG	GCAACCAGTT	TCAGAAGGCT	GAGACCATCC	CTGTGCTGCA	CGAGATGATC	180
CAGCAGATCT	TTAACCTGTT	TAGCACCAAG	GATAGCAGCG	CTGCTTGGGA	TGAGACCCTG	240
CTGGATAAGT	TTTACACCGA	GCTGTACCAG	CAGCTGAACG	ATCTGGAGGC	TTGCGTGATC	300
CAGGGCGTGG	GCGTGACCGA	GACCCCTCTG	ATGAAGGAGG	ATAGCATCCT	GGCTGTGAGG	360
AAGTACTTTC	AGAGGATCAC	CCTGTACCTG	AAGGAGAAGA	AGTACAGCCC	CTGCGCTTGG	420
GAAGTCGTGA	GGGCTGAGAT	CATGAGGAGC	TTTAGCCTGA	GCACCAACCT	GCAAGAGAGC	480
TTGAGGTCTA	AGGAGTAA	498				

Fig. 4

SEQ ID NO: 67

TGCCGCCTTC	TTTGATATTC	ACTCTGTTGT	ATTTTCATCTC	TTCTTGCCGA	TGAAAGGATA	60
TAACAGTCTG	TATAACAGTC	TGTGAGGAAA	TACTTGGTAT	TTCTTCTGAT	CAGTGTTTTT	120
ATAAGTAATG	TTGAATATTG	GATAAGGCTG	TGTGTCCTTT	GTCTTGGGAG	ACAAAGCCCA	180
CAGCAGGTGG	TGGTTGGGGT	GGTGGCAGCT	CAGTGACAGG	AGAGGTTTTT	TTGCCTGTTT	240
TTTTTTTTTT	TTTTTTTTTT	AAGTAAGGTG	TTCTTTTTTC	TTAGTAAATT	TTCTACTGGA	300
CTGTATGTTT	TGACAGGTCA	GAAACATTTT	TTCAAAAGAA	GAACCTTTTG	GAAACTGTAC	360
AGCCCTTTTC	TTTCATTCCC	TTTTTGCTTT	CTGTGCCAAT	GCCTTTGGTT	CTGATTGCAT	420
TATGGAAAAC	GTTGATCGGA	ACTTGAGGTT	TTTATTTATA	GTGTGGCTTG	AAAGCTTGGA	480
TAGCTGTTGT	TACACGAGAT	ACCTTATTAA	GTTTAGGCCA	GCTTGATGCT	TTATTTTTTC	540
CCTTTGAAGT	AGTGAGCGTT	CTCTGGTTTT	TTTCCTTTGA	AACTGGTGAG	GCTTAGATTT	600
TTCTAATGGG	ATTTTTTACC	TGATGATCTA	GTTGCATACC	CAAATGCTTG	TAAATGTTTT	660
CCTAGTTAAC	ATGTTGATAA	CTTCGGATTT	ACATGTTGTA	TATACTTGTC	ATCTGTGTTT	720
CTAGTAAAAA	TATATGGCAT	TTATAGAAAT	ACGTAATTCC	TGATTTCCCT	TTTTTTTATC	780
TCTATGCTCT	GTGTGTACAG	GTCAAACAGA	CTTCACTCCT	ATTTTTATTT	ATAGAATTTT	840
ATATGCAGTC	TGTCGTTGGT	TCTTGTGTTG	TAAGGATACA	GCCTTAAATT	TCCTAGAGCG	900
ATGCTCAGTA	AGGCGGGTTG	TCACATGGGT	TCAAATGTAA	AACGGGCACG	TTTGGCTGCT	960
GCCTTCCCGA	GATCCAGGAC	ACTAAACTGC	TCTGCACATG	AGGTATAAAT	CGCTTCAGAT	1020
CCCAGGGAAG	TGCAGATCCA	CGTGCATATT	CTTAAAGAAG	AATGAATACT	TTCTAAAATA	1080
TTTTGGCATA	GGAAGCAAGC	TGCATGGATT	TGTTTGGGAC	TTAAATTATT	TTGGTAACGG	1140
AGTGCATAGG	TTTTAAACAC	AGTTGCAGCA	TGCTAACGAG	TCACAGCGTT	TATGCAGAAG	1200
TGATGCCTGG	ATGCCTGTTG	CAGCTGTTTA	CGGCACTGCC	TTGCAGTGAG	CATTGCAGAT	1260
AGGGGTGGGG	TGCTTTGTGT	CGTGTTCCCA	CACGCTGCCA	CACAGCCACC	TCCCGGAACA	1320
CATCTCACCT	GCTGGGTACT	TTTCAAACCA	TCTTAGCAGT	AGTAGATGAG	TTACTATGAA	1380
ACAGAGAAGT	TCCTCAGTTG	GATATTCTCA	TGGGATGTCT	TTTTTCCCAT	GTTGGGCCAAA	1440
GTATGATAAA	GCATCTCTAT	TTGTAAATTA	TGCACCTGTT	AGTTCCTGAA	TCCTTTCTAT	1500
AGCACCACTT	ATTGCAGCAG	GTGTAGGCTC	TGGTGTGGCC	TGTGTCTGTG	CTTCAATCTT	1560
TTAAAGCTTC	TTTGGAATAA	CACTGACTTG	ATTGAAGTCT	CTTGAAGATA	GTAAACAGTA	1620
CTTACCTTTG	ATCCCAATGA	AATCGAGCAT	TTCACTTGTA	AAAGAATTCC	GCCTATTCAT	1680
ACCATCTAAT	GTAATTTTAC	ACCCCCAGTG	CTGACACTTT	GGAATATATT	CAAGTAATAG	1740
ACTTTGGCCT	CACCTCTTGG	TGTACTGTAT	TTTGTAATAG	AAAATATTTT	AAACTGTGCA	1800
TATGATTATT	ACATTATGAA	AGAGACATTC	TGCTGATCTT	CAAATGTAAG	AAAATGAGGA	1860
GTGCGTGTGC	TTTTATAAAT	ACAAGTGATT	GCAAATTAGT	GCAGGTGTCC	TTAAAAAATA	1920
AAAAAAAAAAG	TAATATAAAA	AGGACCAGGT	GTTTTACAAG	TGAAATACAT	TCCTATTTGG	1980
TAAACAGTTA	CATTTTTTATG	AAGATTACCA	GCGCTGCTGA	CTTTCTAAAC	ATAAGGCTGT	2040
ATTGTCTTCC	TGTACCATTG	CATTTCCCTCA	TTCCCAATTT	GCACAAGGAT	GTCTGGGTAA	2100
ACTATTCAAG	AAATGGCTTT	GAAATACAGC	ATGGGAGCTT	GTCTGAGTTG	GAATGCAGAG	2160
TTGCACTGCA	AAATGTCAGG	AAATGGATGT	CTCTCAGAAT	GCCCAACTCC	AAAGGATTTT	2220
ATATGTGTAT	ATAGTAAGCA	GTTTCCTGAT	TCCAGCAGGC	CAAAGAGTCT	GCTGAATGTT	2280
GTGTTGCCGG	AGACCTGTAT	TTCTCAACAA	GGTAAGATGG	TATCCTAGCA	ACTGCGGATT	2340
TTAATACATT	TTCAGCAGAA	GTACTTAGTT	AATCTCTACC	TTAGGGATC	GTTTCATCAT	2400
TTTTAGATGT	TATACTTGAA	ATACTGCATA	ACTTTTAGCT	TTCATGGGTT	CCTTTTTTTC	2460
AGCCTTTAGG	AGACTGTAA	GCAATTTGCT	GTCCAACTTT	TGTGTTGGTC	TTAAACTGCA	2520
ATAGTAGTTT	ACCTTGATTT	GAAGAAATAA	AGACCATTTT	TATATTAAAA	AATACTTTTG	2580
TCTGTCTTCA	TTTTGACTTG	TCTGATATCC	TTGCAGTGCC	CATTATGTCA	GTTCTGTCAG	2640
ATATTCAGAC	ATCAAACTT	AACGTGAGCT	CAGTGGAGTT	ACAGCTGCGG	TTTTGATGCT	2700
GTTATTATTT	CTGAAACTAG	AAATGATGTT	GTCTTCATCT	GCTCATCAAA	CACTTCATGC	2760
AGAGTGTAAG	GCTAGTGAGA	AATGCATACA	TTTATTGATA	CTTTTTTAAA	GTCAACTTTT	2820
TATCAGATTT	TTTTTTCATT	TGGAAATATA	TTGTTTTCTA	GACTGCATAG	CTTCTGAATC	2880
TGAAATGCAG	TCTGATTGGC	ATGAAGAAGC	ACAGCACTCT	TCATCTTACT	TAAACTTCAT	2940
TTTGGAATGA	AGGAAGTTAA	GCAAGGGCAC	AGGTCCATGA	AATAGAGACA	GTGCGCTCAG	3000
GAGAAAGTGA	ACCTGGATTT	CTTTGGCTAG	TGTTCTAAAT	CTGTAGTGAG	GAAAGTAACA	3060
CCCGATTCCCT	TGAAAGGGCT	CCAGCTTTAA	TGCTTCCAAA	TTGAAGGTGG	CAGGCAACTT	3120
GGCCACTGGT	TATTTACTGC	ATTATGTCTC	AGTTTCGCAG	CTAACCTGGC	TTCTCCACTA	3180

TTGAGCATGG	ACTATAGCCT	GGCTTCAGAG	GCCAGGTGAA	GGTTGGGATG	GGTGGAAGGA	3240
GTGCTGGGCT	GTGGCTGGGG	GGACTGTGGG	GACTCCAAGC	TGAGCTTGGG	GTGGGCAGCA	3300
CAGGGAAAAG	TGTGGGTAAC	TATTTTAAAG	TACTGTGTTG	CAAACGTCTC	ATCTGCAAAT	3360
ACGTAGGGTG	TGTACTCTCG	AAGATTAACA	GTGTGGGTTC	AGTAATATAT	GGATGAATTC	3420
ACAGTGGAAG	CATTCAAGGG	TAGATCATCT	AACGACACCA	GATCATCAAG	CTATGATTGG	3480
AAGCGGTATC	AGAAGAGCGA	GGAAGGTAAG	CAGTCTTCAT	ATGTTTTCCC	TCCACGTAAA	3540
GCAGTCTGGG	AAAGTAGCAC	CCCTTGAGCA	GAGACAAGGA	AATAATTGAG	GAGCATGTGC	3600
TAGGAGAACT	TTCTTGCTGA	ATTCTACTTG	CAAGAGCTTT	GATGCCTGGC	TTCTGGTGCC	3660
TTCTGCAGCA	CCTGCAAGGC	CCAGAGCCTG	TGGTGAGCTG	GAGGGAAAGA	TTCTGCTCAA	3720
GTCCAAGCTT	CAGCAGGTCA	TTGTCTTTGC	TTCTTCCCCC	AGCACTGTGC	AGCAGAGTGG	3780
AACTGATGTC	GAAGCCTCCT	GTCCACTACC	TGTTGCTGCA	GGCAGACTGC	TCTCAGAAAA	3840
AGAGAGCTAA	CTCTATGCCA	TAGTCTGAAG	GTAAAATGGG	TTTAAAAAAA	GAAAACACAA	3900
AGGCAAAACC	GGCTGCCCCA	TGAGAAGAAA	GCAGTGGTAA	ACATGGTAGA	AAAGGTGCAG	3960
AAGCCCCCAG	GCAGTGTGAC	AGGCCCTCC	TGCCACCTAG	AGGCGGGAAC	AGGCTTCCCT	4020
GCCTAGGGCT	CTGCCCGCGA	AGTGCCTGTT	TCTTTGGTGG	GTTTTGTTTG	GCGTTTGGTT	4080
TTGAGATTTA	GACACAAGGG	AAGCCTGAAA	GGAGGTGTTG	GGCACTATTT	TGGTTTGTA	4140
AGCCTGTACT	TCAAATATAT	ATTTTGTGAG	GGAGTGTAGC	GAATTGGCCA	ATTTAAAAATA	4200
AAGTTGCAAG	AGATTGAAGG	CTGAGTAGTT	GAGAGGGTAA	CACGTTTAAT	GAGATCTTCT	4260
GAAACTACTG	CTTCTAAACA	CTTGTTTGAG	TGGTGAGACC	TTGGATAGGT	GAGTGCTCTT	4320
GTTACATGTC	TGATGCACTT	GCTTGTCTTT	TTCCATCCAC	ATCCATGCAT	TCCACATCCA	4380
CGCATTGTGTC	ACTTATCCCA	TATCTGTGTCAT	ATCTGACATA	CCTGTCTCTT	CGTCACTTGG	4440
TCAGAAGAAA	CAGATGTGAT	AATCCCCAGC	CGCCCCAAGT	TTGAGAAGAT	GGCAGTTGCT	4500
TCTTTCCCTT	TTTCCTGCTA	AGTAAGGATT	TTCTCCTGGC	TTTGACACCT	CACGAAATAG	4560
TCTTCCTGCC	TTACATTCTG	GGCATTATTT	CAAATATCTT	TGGAGTGCGC	TGCTCTCAAG	4620
TTTGTGTCTT	CCTACTCTTA	GAGTGAATGC	TCTTAGAGTG	AAAGAGAAGG	AAGAGAAGAT	4680
GTTGGCCGCA	GTTCTCTGAT	GAACACACCT	CTGAATAATG	GCCAAAGGTG	GGTGGGTTTC	4740
TCTGAGGAAC	GGGCAGCGTT	TGCCCTTGAA	AGCAAGGAGC	TCTGCGGAGT	TGCAGTTATT	4800
TTGCAACTGA	TGGTGGAAC	GGTGCTTAAA	GCAGATTCCC	TAGGTTCCCT	GCTACTTCTT	4860
TTCTTCTTGG	GCAGTCAGTT	TATTTCTGAC	AGACAAACAG	CCACCCCCAC	TGCAGGCTTA	4920
GAAAGTATGT	GGCTCTGCC	GGGTGTGTTA	CAGCTCTGCC	CTGGTGAAAG	GGGATTAAAA	4980
CGGGCACCAT	TCATCCCCAA	CAGGATCCTC	ATTGATGGAT	CAAGCTGTAA	GGAAGTTGGG	5040
CTCCAACCTC	AAAACATTAA	TTGGAGTACG	AATGTAATTA	AAACTGCATT	CTCGCATTC	5100
TAAGTCATTT	AGTCTGGACT	CTGCAGCATG	TAGGTCGGCA	GCTCCCACTT	TCTCAAAGAC	5160
CACTGATGGA	GGAGTAGTAA	AAATGGAGAC	CGATTGAGAA	CAACCAACGG	AGTGTGGCCG	5220
AAGAAACTGA	TGGAAATAAT	GCATGAATTG	TGTGGTGGAC	ATTTTTTTT	AATACATAAA	5280
CTACTTCAAA	TGAGGTCGGA	GAAGGTCAGT	GTTTTATTAG	CAGCCATAAA	ACCAGGTGAG	5340
CGAGTACCAT	TTTTCTCTAC	AAGAAAAACG	ATTCTGAGCT	CTGCGTAAGT	ATAAGTTCTC	5400
CATAGCGGCT	GAAGCTCCCC	CCTGGCTGCC	TGCCATCTCA	GCTGGAGTGC	AGTGCCATTT	5460
CCTTGGGGTT	TCTCTCACAG	CAGTAATGGG	ACAATACTTC	ACAAAAATTC	TTTCTTTTCC	5520
TGTCATGTGG	GATCCCTACT	GTGCCCTCCT	GGTTTTACGT	TACCCCTGTA	CTGTTCCATT	5580
CAGCGGTTTG	GAAAAGAGAA	AAGAATTTGG	AAATAAAACA	TGTCTACGTT	ATCACCTCCT	5640
CCAGCATTTT	GGTTTTTAAT	TATGTCAATA	ACTGGCTTAG	ATTTGGAAAT	GAGAGGGGGT	5700
TGGGTGTATT	ACCGAGGAAC	AAAGGAAGGC	TTATATAAAC	TCAAGTCTTT	TATTTAGAGA	5760
ACTGGCAAGC	TGTCAAAAAC	AAAAAGGCCT	TACCACCAAA	TTAAGTGAAT	AGCCGCTATA	5820
GCCAGCAGGG	CCAGCACGAG	GGATGGTGCA	CTGCTGGCAC	TATGCCACGG	CCTGCTTGTTG	5880
ACTCTGAGAG	CAACTGCTTT	GGAAATGACA	GCACTTGGTG	CAATTTCTTT	TGTTTCAGAA	5940
TGCGTAGAGC	GTGTGCTTGG	CGACAGTTTT	TCTAGTTAGG	CCACTTCTTT	TTTCCTTCTC	6000
TCCTCATTCT	CCTAAGCATG	TCTCCATGCT	GGTAATCCCA	GTCAAGTGAA	CGTTCAAACA	6060
ATGAATCCAT	CACTGTAGGA	TTCTCGTGGT	GATCAAATCT	TTGTGTGAGG	TCTATAAAAT	6120
ATGGAAGCTT	ATTTATTTTT	CGTTCTTCCA	TATCAGTCTT	CTCTATGACA	ATTCACATCC	6180
ACCACAGCAA	ATTAAAGGTG	AAGGAGGCTG	GTGGGATGAA	GAGGGTCTTC	TAGCTTTACG	6240
TTCTTCCTTG	CAAGGCCACA	GGAAATGCTG	GAGAGCTGTA	GAATACAGCC	TGGGGTAAGA	6300
AGTTCAGTCT	CCTGCTGGGA	CAGCTAACCG	CATCTTATAA	CCCCTTCTGA	GACTCATCTT	6360
AGGACCAAAT	AGGGTCTATC	TGGGGTTTTT	GTTCTGTCTG	TTCTCTCTGG	AAGGCTATCT	6420
CACTATTTCA	CTGCTCCAC	GGTTACAAAC	CAAAGATACA	GCCTGAATTT	TTTCTAGGCC	6480

ACATTACATA	AATTTGACCT	GGTACCAATA	TTGTTCTCTA	TATAGTTATT	TCCTTCCCCA	6540
CTGTGTTTAA	CCCTTAAAGG	CATTTCAGAAC	AAGTAGAATC	ATAGAATGGT	TTGGATTGGA	6600
AGGGGCTTAA	AACATCATCC	ATTTCCAACC	CTCTGCCATG	GGCTGCTTGC	CACCCACTGG	6660
CTCAGGCTGC	CCAGGGCCCC	ATCCAGCCTG	GCCTTGAGCA	CCTCCAGGGA	TGGGGCACCC	6720
ACAGCTTCTC	TGGGCAGCCT	GTGCCAACAC	CTCACCCTC	TCTGGGTAAA	GAATTCTCTT	6780
TTAACATCTA	ATCTAAATCT	CTTCTCTTTT	AGTTTAAAGC	CATTCTCTTT	TTTCCCGTTG	6840
CTATCTGTCC	AAGAAATGTG	TATTGGTCTC	CCTCCTGCTT	ATAAGCAGGA	AGTACTGGAA	6900
GGCTGCAGTG	AGGTCTCCCC	ACAGCCTTCT	CTTCTCCAGG	CTGAACAAGC	CCAGCTCCTT	6960
CAGCCTGTCT	TCGTAGGAGA	TCATCTTAGT	GGCCCTCCTC	TGGACCCATT	CCAACAGTTC	7020
CACGGCTTTC	TTGTGGAGCC	CCAGGTCTGG	ATGCAGTACT	TCAGATGGGG	CCTTACAAAG	7080
GCAGAGCAGA	TGGGGACAAT	CGCTTACCCC	TCCCTGCTGG	CTGCCCCGTG	TTTGATGCAG	7140
CCCAGGGTAC	TGTTGGCCTT	TCAGGCTCCC	AGACCCCTTG	CTGATTTGTG	TCAAGCTTTT	7200
CATCCACCAG	AACCCACGCT	TCCTGGTTAA	TACTTCTGCC	CTCACTTCTG	TAAGCTTGT	7260
TCAGGAGACT	TCCATTCTTT	AGGACAGACT	TGTGTACACC	TACCTGCCCT	ATTCTTGCA	7320
ATATACATTT	CAGTTCATGT	TTCCTGTAAC	AGGACAGAAT	ATGTATTCCCT	CTAACAAAAA	7380
TACATGCAGA	ATTCTTAGTG	CCATCTCAGT	AGGGTTTTCA	TGGCAGTATT	AGCACATAGT	7440
CAATTTGCTG	CAAGTACCTT	CCAAGCTGCG	GCCTCCCAT	AATCCTGTAT	TTGGGATCAG	7500
TTACCTTTTG	GGGTAAGCTT	TTGTATCTGC	AGAGACCCTG	GGGGTTCTGA	TGTGCTTCAG	7560
CTCTGCTCTG	TTCTGACTGC	ACCATTTTCT	AGATCACCCA	GTTGTTCCCTG	TACAACCTCC	7620
TTGTCCTCCA	TCCTTTCCCA	GCTTGTATCT	TTGACAAATA	CAGGCCTATT	TTTGTGTTTG	7680
CTTCAGCAGC	CATTTAATTC	TTCAGTGTCA	TCTTGTTCTG	TTGATGCCAC	TGGAACAGGA	7740
TTTTCAGCAG	TCTTGCAAAG	AACATCTAGC	TGAAAACCTT	CTGCCATTCA	ATATTCTTAC	7800
CAGTTCTTCT	TGTTTGAGGT	GAGCCATAAA	TTACTAGAAC	TTCGTCACTG	ACAAGTTTAT	7860
GCATTTTATT	ACTTCTATTA	TGTACTTACT	TTGACATAAC	ACAGACACGC	ACATATTTTG	7920
CTGGGATTTT	CACAGTGTCT	CTGTGTCTCT	CACATGGTTT	TACTGTCATA	CTTCCGTTAT	7980
AACCTTGCCA	ATCTGCCCAG	CTGCCCATCA	CAAGAAAAGA	GATTCCCTTT	TTATTACTTC	8040
TCTTCAGCCA	ATAAACAAAA	TGTGAGAAGC	CCAAACAAGA	ACTTGTGGGG	CAGGCTGCCA	8100
TCAAGGGAGA	GACAGCTGAA	GGGTTGTGTA	GCTCAATAGA	ATTAAGAAAT	AATAAAGCTG	8160
TGTCAGACAG	TTTTGCCTGA	TTTATACAGG	CACGCCCCAA	GCCAGAGAGG	CTGTCTGCCA	8220
AGGCCACCTT	GCAGTCCTTG	GTTTGTAAGA	TAAGTCATAG	GTAACCTTTC	TGGTGAATTG	8280
CGTGGAGAAT	CATGATGGCA	GTTCTTGCTG	TTTACTATGG	TAAGATGCTA	AAATAGGAGA	8340
CAGCAAAGTA	ACACTTGCTG	CTGTAGGTGC	TCTGCTATCC	AGACAGCGAT	GGCACTCGCA	8400
CACCAAGATG	AGGGATGCTC	CCAGCTGACG	GATGCTGGGG	CAGTAACAGT	GGGTCCCATG	8460
CTGCCTGCTC	ATTAGCATCA	CCTCAGCCCT	CACCAGCCCA	TCAGAAGGAT	CATCCCAAGC	8520
TGAGGAAAGT	TGCTCATCTT	CTTCACATCA	TCAAACCTTT	GGCCTGACTG	ATGCCTCCCG	8580
GATGCTTAAA	TGTGGTCACT	GACATCTTTA	TTTTTCTATG	ATTTCAAGTC	AGAACCTCCG	8640
GATCAGGAGG	GAACACATAG	TGGGAATGTA	CCCTCAGCTC	CAAGGCCAGA	TCTTCCTTCA	8700
ATGATCATGC	ATGCTACTTA	GGAAGGTGTG	TGTGTGTGAA	TGTAGAATTG	CCTTTGTTAT	8760
TTTTTCTTCC	TGCTGTCAAG	AACATTTTGA	ATACCAGAGA	AAAAGAAAAG	TGCTCTTCTT	8820
GGCATGGGAG	GAGTTGTCAC	ACTTGCAAAA	TAAAGGATGC	AGTCCCAAAT	GTTTATAATC	8880
TCAGGGTCTG	AAGGAGGATC	AGAAACTGTG	TATACAATTT	CAGGCTTCTC	TGAATGCAGC	8940
TTTTGAAAGC	TGTTCTTGGC	CGAGGCAGTA	CTAGTCAGAA	CCCTCGGAAA	CAGGAACAAA	9000
TGTCTTCAAG	GTGCAGCAGG	AGGAAACACC	TTGCCCATCA	TGAAAGTGAA	TAACCACTGC	9060
CGCTGAAGGA	ATCCAGCTCC	TGTTTGAGCA	GGTGCTGCAC	ACTCCACAC	TGAAACAACA	9120
GTTTCAATTTT	ATAGGACTTC	CAGGAAGGAT	CTTCTTCTTA	AGCTTCTTAA	TTATGGTACA	9180
TCTCCAGTTG	GCAGATGACT	ATGACTACTG	ACAGGAGAAT	GAGGAAGTAG	CTGGGAATAT	9240
TTCTGTTTGA	CCACCATGGA	GTCACCCATT	TCTTTACTGG	TATTTGGAAA	TAATAATTCT	9300
GAATTGCAAA	GCAGGAGTTA	GCGAAGATCT	TCATTTCTTC	CATGTTGGTG	ACAGCACAGT	9360
TCTGGCTATG	AAAGTCTGCT	TACAAGGAAG	AGGATAAAAA	TCATAGGGAT	AATAAATCTA	9420
AGTTTGAAGA	CAATGAGGTT	TTAGCTGCAT	TTGACATGAA	GAAATTGAGA	CCTCTACTGG	9480
ATAGCTATGG	TATTTACGTG	TCTTTTTGCT	TAGTTACTTA	TTGACCCAG	CTGAGGTCAA	9540
GTATGAATCT	AGGCTCTCTG	GGCTACTGGC	ATGGATTGAT	TACATACAAC	TGTAATTTTA	9600
GCAGTGAATTT	AGGGTTTATG	AGTACTTTTG	CAGTAAATCA	TAGGGTTAGT	AATGTTAATC	9660
TCAGGGAAAA	AAAAAAAAAAG	CCAACCCTGA	CAGACATCCC	AGCTCAGGTG	GAAATCAAGG	9720
ATCACAGCTC	AGTGCGGTCC	CAGAGAACAC	AGGGACTCTT	CTCTTAGGAC	CTTTATGTAC	9780

```

AGGGCCTCAA GATAACTGAT GTTAGTCAGA AGACTTTCCA TTCTGGCCAC AGTTCAGCTG 9840
AGGCAATCCT GGAATTTTCT CTCCGCTGCA CAGTTCCAGT CATCCAGTT TGTACAGTTC 9900
TGGCACTTTT TGGGTCAGGC CGTGATCCAA GGAGCAGAAG TTCCAGCTAT GGTACAGGAG 9960
TGCCTGACCG TCCCAACTCA CTGCACTCAA ACAAAGGCGA AACCACAAGA GTGGCTTTTG 10020
TTGAAATTGC AGTGTGGCCC AGAGGGGCTG CACCAGTACT GGATTGACCA CGAGGCAACA 10080
TTAATCCTCA GCAAGTGCAA TTTGCAGCCA TTAAATTGAA CTAAGTGATA CTACAATGCA 10140
ATCAGTATCA ACAAGTGGTT TGGCTTGGA GATGGAGTCT AGGGGCTCTA CAGGAGTAGC 10200
TACTCTCTAA TGGAGTTGCA TTTTGAAGCA GGACACTGTG AAAAGCTGGC CTCCTAAAGA 10260
GGCTGCTAAA CATTAGGGTC AATTTTCCAG TGCACCTTCT GAAGTGTCTG CAGTTCCCCA 10320
TGCAAAGCTG CCCAAACATA GCACTTCCAA TTGAATACAA TTATATGCAG GCGTACTGCT 10380
TCTTGCCAGC ACTGTCCTTC TCAAATGAAC TCAACAAACA ATTTCAAAGT CTAGTAGAAA 10440
GTAACAAGCT TTGAATGTCA TTA AAAAAGTA TATCTGCTTT CAGTAGTTCA GCTTATTTAT 10500
GCCCCACTAGA AACATCTTGT ACAAGCTGAA CACTGGGGCT CCAGATTAGT GGTAAACCT 10560
ACTTTATACA ATCATAGAAT CATAGAATGG CCTGGGTTGG AAGGGACCCC AAGGATCATG 10620
AAGATCCAAC ACCCCCGCCA CAGGCAGGGC CACCAACCTC CAGATCTGGT ACTAGACCAG 10680
GCAGCCCAGG GCTCCATCCA ACCTGGCCAT GAACACCTCC AGGGATGGAG CATCCACAAC 10740
CTCTCTGGGC AGCCTGTGCC AGCACCTCAC CACCCTCTCT GTGAAGAACT TTTCCCTGAC 10800
ATCCAATCTA AGCCTTCCCT CCTTGAGGTT AGATCCACTC CCCCTTGTGC TATCACTGTC 10860
TACTCTTGTA AAAAGTTGAT TCTCCTCCTT TTTGGAAGGT TGCAATGAGG TCTCCTTGCA 10920
GCCTTCTTCT CTCTGTCAGG ATGAACAAGC CCAGCTCCCT CAGCCTGTCT TTATAGGAGA 10980
GGTGCTCCAG CCCTCTGATC ATCTTTGTGG CCCTCCTCTG GACCCGCTCC AAGAGCTCCA 11040
CATCTTTTCT GTACTGGGGG CCCCAGGCCT GAATGCAGTA CTCCAGATGG GGCCTCAAAA 11100
GAGCAGAGTA AAGAGGGACA ATCACCTTCC TCACCCTGCT GGCCAGCCCT CTTCTGATGG 11160
AGCCCTGGAT ACAACTGGCT TTCTGAGCTG CAACTTCTCC TTATCAGTTC CACTATTAAA 11220
ACAGGAACAA TACAACAGGT GCTGATGGCC AGTGCAGAGT TTTTCACACT TCTTCATTTT 11280
GGTAGATCTT AGATGAGGAA CGTTGAAGTT GTGCTTCTGC GTGTGCTTCT TCCTCCTCAA 11340
ATACTCCTGC CTGATACCTC ACCCCACCTG CCACTGAATG GCTCCATGGC CCCCTGCAGC 11400
CAGGGCCCTG ATGAACCCGG CACTGCTTCA GATGCTGTTT AATAGCACAG TATGACCAAG 11460
TTGCACCTAT GAATACACAA ACAATGTGTT GCATCCTTCA GCACTTGAGA AGAAGAGCCA 11520
AATTTGCATT GTCAGGAAAT GGTTTAGTAA TTCTGCCAAT TAAAACTTGT TTATCTACCA 11580
TGGCTGTTTT TATGGCTGTT AGTAGTGGTA CACTGATGAT GAACAATGGC TATGCAGTAA 11640
AATCAAGACT GTAGATATTG CAACAGACTA TAAAATTCCT CTGTGGCTTA GCCAATGTGG 11700
TACTTCCCAC ATTGTATAAG AAATTTGGCA AGTTTAGAGC AATGTTTGAA GTGTTGGGAA 11760
ATTTCTGTAT ACTCAAGAGG GCGTTTTTGA CAACTGTAGA ACAGAGGAAT CAAAAGGGGG 11820
TGGGAGGAAG TTAAAAGAAG AGGCAGGTGC AAGAGAGCTT GCAGTCCCGC TGTGTGTACG 11880
ACACTGGCAA CATGAGGTCT TTGCTAATCT TGGTGCTTTG CTTCTGCCC CTGGCTGCC 11940
TAGGG 11945

```

FIG. 5

SEQ ID NO: 68

```
AAAGTCTAGAGTCGGGGCGGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAG 60
TTTGGACAAACCACAAC TAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGAT 120
GCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAACAAGTTAACAACAACAATTGC 180
ATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTTAAAGCAAGTAAAC 240
CTCTACAAATGTGGTAAAATCGATAAGGATCCGTCGAGCGGCCGC 285
```

FIG. 6

SEQ ID NO: 69

1 CGCGTGGTAGGTGGCGGGGGGTTCCAGGAGAGCCCCAGCGCGGACGGC
AGCGCCGTCACTCACCCTCCGTCTCCCTCCGCCCAGGGTCGCCTGGCGC
AACCCTGCAAGGGCACCAGCTCCAGGCGTGGATCAGAGGCTGCCGGCT
GTGAGGAGCTGCCGCGCCCGCCCGCTGCACAGCCGGCCGCTTTGC
200 GAGCGCGACGCTACCCGCTTGGCAGTTTTAAACGCATCCCTCATTAAC
GACTATACGCAAACGCCTTCCCGTCGGTCCGCGTCTCTTTCCGCCGCCAG
GGCGACACTCGCGGGGAGGGCGGGAAGGGGGCCGGGCGGGAGCCCGCGGC
CAACCGTCGCCCCGTGACGGCACCGCCCCGCCCCGTGACGCGGTGCGGG
400 CGCCGGGGCCGTGGGGCTGAGCGCTGCGGCGGGGCGGGCCGGGCGGGG
CGGGAGCTGAGCGCGGCGCGGCTGCGGGCGGGCGCCCCCTCCGGTGCAATA
TGTTCAAGAGAATGGCTGAGTTCGGGCTGACTCCGGGGCAGGGTGAAG
GTGCGGCGCGGGCGGAGGGACGGGGCGGGCGCGGGGCCGCCGGCGGGTG
600 CCGGGGCTCTGCGGGCCCGCCCGGCTCGGGCTGCTGCGGCGCTTACGGG
CGCGCTTCTCGCCGCTGCCGCTTCTCTTCTCTCCCGCGCAAGGGCGTCAC
CATCGTGAAGCCGGTAGTGTACGGGAACGTGGCGCGGTACTTCGGGAAGA
AGAGGGAGGAGGACGGGCACACGCATCAGTGGACGGTTTACGTGAAGCCC
800 TACAGGAACGAGGTAGGGCCCCGAGCGCGTCGGCCGCCGTTCTCGGAGCGC
CGGAGCCGTCAGCGCCGCGCCTGGGTGCGCTGTGGGACACAGCGAGCTTC
TCTCGTAGGACATGTCCGCCTACGTGAAAAAATCCAGTTCAAGCTGCAC
GAGAGCTACGGGAATCCTCTCCGAGGTGGGTGTTGCGTCGGGGGGTTTGC
1000 TCCGCTCGGTCCCGCTGAGGCTCGTCGCCCTCATCTTTCTTTCTGTCGCGC
AGTCGTTACCAAACCGCCGTACGAGATCACCGAAACGGGCTGGGGCGAAT
TTGAAATCATCATCAAGATATTTTTCATTGATCCAAACGAGCGACCCGTA
AGTACGCTCAGCTTCTCGTAGTGCTTCCCCCGTCCCTGGCGGGCCGGGGCT
1200 GGGCTGCTCGCTGCTGCCGGTCACAGTCCCGCCAGCCGCGGAGCTGACTG
AGCTCCCTTTCCCGGGACGTGTGCTCTGTGTTTCGGTCAGCGAGGCTATCG
GGAGGGCTTTGGCTGCATTTGGCTTCTCTGGCGCTTAGCGCAGGAGCACG
TTGTGCTACGCCTGAACCTACAGCTGTGAGAAGGCCGTGGAAACCGCTCTC
1400 AAAGTATTTATTTGGCGAAATGGCTCTAAACTAAATCGTCTCCTCTCTTT
GGAAATGCTTTAGAGAAGGTCTCTGTGGTAGTTCTTATGCATCTATCCTA
AAGCACTTGGCCAGACAATTTAAAGACATCAAGCAGCATTTATAGCAGGC
ACGTTTAAATAACGAATACTGAATTTAAGTAACTCTGCTCACGTTGTATGA
1600 CGTTTATTTTCGTATTCTGAAAGCCATTAAAACTCTGTGCAGTTGTTTA
GTAAGAACAGCTGCCACTGTTTTGTATCTAGGAGATAACTGGTGTTTCCC
TACAGTTCTCAAGCTGATAAACTCTGTCTTTGTATCTAGGTAACCCGTGT
ATCACTTGCTGAAGCTTTTTTCAGTCTGACACCAATGCAATCCTGGGAAAG
1800 AAACTGTAGTTTCTGAATTCATGATGAAATGGTATGAAAATTTTAATG
TCAACCGAGCCTGACTTTATTTAAAAAAATTTATTGATGGTGCTGTGTAT
TTTGGTCCCTTCCCTAGATATTTCAAGATCCTACTGCCATGATGCAGCAAC
TGCTAACGACGTCCCGTCAGCTGACACTTGGTGCTTACAAGCATGAAACA
2000 GAGTGTAAGTGCAAAATGAGGATACCTTCGCCGACCGTCATTCACTACTA
ATGTTTTCTGTGGGATGTGATCGTACAGTGAGTTTGGCTGTGTGAAATTT
GAATAGCTTGGTATTGGCAGTGATGACGTGATCGATGCCTTGCTTATCAT
GTTTGAAATGAAGTAGAATAAATGCAGCCTGCTTTATTTGAGATAGTTTG
2200 GTTCATTTTATGGAATGCAAGCAAGATTATACTTCCTCACTGAATTGCA
CTGTCCAAAGGTGTGAAATGTGTGGGGATCTGGAGGACCGTGACCGAGGG
ACATTGGATCGCTATCTCCCATTTCTTTTGTGTTACCAGTTTCAAGATTTT

CTTTTCACCTAGTCTTTAATTCCCAGGGTTTTGTTTTTTCCTTGGTCATA
2400 GTTTTTGTTTTTTCACTCTGGCAAATGATGTTGTGAATTACACTGCTTCAG
CCACAAAACATGATGGACTGAATGAGGTCATCAAACAACTTTTCTTCTTC
CGTATTTCCCTTTTTTTTTCCCCCACTTATCATTTTTTACTGCTGTTGTTGAG
TCTGTAAGGCTAAAAGTAACTGTTTTGTGCTTTTTTCAGGACGTGTGCTTT
2600 CCAAATTACTGCCACATATATAAAGAAAGGTTGGAATTTTAAAGATAATT
CATGTTTCTTCTTCTTTTTTGGCCACCACAGTTGCAGATCTTGAAGTAAAA
ACCAGGGAAAAGCTGGAAGCTGCCAAAAAGAAAACCAGTTTGAATTTGC
TGAGCTTAAAGAAAGGTTAAAAGCAAGTCGTGAAACCATCAACTGCTTAA
2800 AGAGTGAAATCAGAAAACCTCGAAGAGGATGATCAGTCTAAAGATATGTGA
TGAGTGTGACTTGGCAGGGAGCCTATAATGAGAATGAAAGGACTTCAGT
CGTGAGGTTGTATGCGTTCTCTCCAATTCTGTAACGGAGACTGTATGAAT
TTCATTTGCAAATCACTGCAGTGTGTGACAACACTGACTTTTTATAAATGGC
3000 AGAAAAACAAGAATGAATGTATCCTCATTTTTATAGTTAAATCTATGGGTA
TGTACTGGTTTATTTCAAGGAGAATGGATCGTAGAGACTTGGAGGCCAGA
TTGCTGCTTGTATTGACTGCATTTGAGTGGTGTAGGAACATTTTGTCTAT
GGTCCCGTGTAGTTTACAGAATGCCACTGTTCACTGTTTTGTTTTGTAT
3200 TTTACTTTTTTCTACTGCAACGTCAAGGTTTTTAAAAGTTGAAAATAAAACA
TGCAGGTTTTTTTTTAAATATTTTTTGTCTCTATCCAGTTTGGGCTTCAA
GTATTATTGTTAACAGCAAGTCCTGATTTAAGTCAGAGGCTGAAGTGTAA
TGGTATTCAAGATGCTTAAGTCTGTTGTCAGCAAAACAAAAGAGAAAAC
3400 TCATAAAATCAGGAAGTTGGCATTCTAATAACTTCTTTATCAACAGATA
AGAGTTTCTAGCCCTGCATCTACTTTCACTTATGTAGTTGATGCCTTTAT
ATTTTGTGTGTTTGGATGCAGGAAGTGATTCCCTACTCTGTTATGTAGATA
TTCTATTTAACACTTGTACTCTGCTGTGCTTAGCCTTTCCCCATGAAAAT
3600 TCAGCGGCTGTAAATCCCCCTCTTCTTTGTAGCCTCATAACAGATGGCAG
ACCTCAGGCTTATAAAGGCTTGGGCATCTTCTTTACTGCTTTGAGATTC
TGTGTTGCAGTAACCTCTGCCAGAGAGGAGAAAAGCCCCACAAACCTCAT
CCCCCTTCTTCTATAGCAATCAGTATTACTAATGCTTTGAGAACAGAGCAC
3800 TGGTTTGAAACGTTTGATAATTAGCATTTAACATGGCTTGGTAAAGATGC
AGAAGTGAACAGCTGTGACAGTATGAACTCAGTATGGAGACTTCATTAA
GACAAACAGCTGTTAAAATCAGGCATGTTTCATTGAGGAGGACGGGGCAA
CTTGACACAGTGGTGCCACACAAATCCTTCCTGGCGCTGCAGACCAATT
4000 TTTCTGGCATTCTGACTGCCGTTGCTGCTGGTCACAGAGAGCAACTATTT
TTATCAGCCACAGGCAATTTGCTTGTAGTATTTTCCAAGTGTGTAGGTA
AGTATAAATGCATCGGCTCCAGAGCACTTTGAGTATACTTATTAAAAACA
TAAATGAAAGACAAATTAGCTTTGCTTGGGTGCACAGAACATTTTTAGTT
4200 CCAGCCTGCTTTTTGGTAGAAGCCCTCTTCTGAGGCTAGAAGTACTTTG
ACAAGTAGAGAACTGGCAACGGAGCTATTGCTATCGAAGGATCCTTGTT
AACAAAGTTAATCGTCTTTTAAGGTTTGGTTTATTCATTAAATTTGCTTT
TAAGCTGTAGCTGAAAAAGAACGTGCTGTCTTCCATGCACCAGGTGGCAG
4400 CTCTGTGCAAAGTGCTCTCTGGTCTCACCAGCCTTTTAATTGCCGGGATT
CTGGCACGTCTGAGAGGGCTCAGACTGGCTTCGTTTGTGTTGAACAGCGTG
TACTGCTTTCTGTAGACATGGCCGGTTTCTCTCCTGCAGCTTATGAACT
GTTACACTGAACACACTGGAACAGGTTGCCCAAGGAGGCCGTGGATGCC
4600 CCATCCCTGGAGGCATTCAAGGCCAGGCTGGATGTGGCTCTGGGCAGCCT
GGTCTGGTGGTTGGCGATCCTGCACATAGCAGCGGGTTGAACTCGATG
ATCACTGTGGTCCTTTTCAACCCAGGCTATTCTATGATTCTATGATTCAA
CAGCAAATCATATGTACTGAGAGAGGAAACAAACACAAGTGCTACTGTTT
4800 GCAAGTTTTGTTCAATTTGGTAAAAGAGTCAGGTTTTTAAATTCAAAATCT

GTCTGGTTTTGGTGTTTTTTTTTTTTTTTATTTATTATTTCTTTGGGGTTCT
TTTTGATGCTTTATCTTTCTCTGCCAGGACTGTGTGACAATGGGAACGAA
AAAGAACATGCCAGGCACTGTCTGGATTGCACACGCTGGTTGCACTCAG
5000 TAGCAGGCTCAGAACTGCCAGTCTTTCCACAGTATTACTTTCTAAACCTA
ATTTTAATAGCGTTAGTAGACTTCCATCACTGGGCAGTGCTTAGTGAATG
CTCTGTGTGAACGTTTTACTTATAAGCATGTTGGAAGTTTTGATGTTCCCT
GGATGCAGTAGGGAAGGACAGATTAGCTATGTGAAAAGTAGATTCTGAGT
5200 ATCGGGGTTACAAAAAGTATAGAAACGATGAGAAATTCCTTGTTGTAACCTA
ATTGGAATTTCTTTAAGCGTTCACCTATGCTACATTCATAGTATTTCCAT
TTAAAAGTAGGAAAAGGTAAAACGTGAAATCGTGTGATTTTCGGATGGAA
CACCGCCTTCCTATGCACCTGACCAACTTCCAGAGGAAAAGCCTATTGAA
5400 AGCCGAGATTAAGCCACCAAAAGAACTCATTTCGATTGGAATATGTAGTA
TTTGCCCTCTTCCTCCCGGGTAATTACTATACTTTATAGGGTGCTTATAT
GTTAAATGAGTGGCTGGCACTTTTTATTCTCACAGCTGTGGGGAATTCTG
TCCTCTAGGACAGAAACAATTTTAATCTGTTCCACTGGTGACTGCTTTGT
5600 CAGCACTTCCACCTGAAGAGATCAATACACTCTTCAATGTCTAGTTCTGC
AACACTTGGCAAACCTCACATCTTATTTTCATACTCTCTTCATGCCTATGC
TTATTAAAGCAATAATCTGGGTAATTTTTTGTTTTAATCACTGTCCTGACC
CCAGTGATGACCGTGTCCACCTAAAGCTCAATTCAGGTCCTGAATCTCT
5800 TCAACTCTCTATAGCTAACATGAAGAATCTTCAAAAAGTTAGGTCTGAGGG
ACTTAAGGCTAACTGTAGATGTTGTTGCCTGGTTTCTGTGCTGAAGGCCG
TGTAGTAGTTAGAGCATTCAACCTCTAG

Fig. 7

SEQ ID NO: 70

1 TGCCGCCTTCTTTGATATTCACCTCTGTTGTATTTTCATCTCTTCTTGCCGA
TGAAAGGATATAACAGTCTGTATAACAGTCTGTGAGGAAATACTTGGTAT
TTCTTCTGATCAGTGTTTTTATAAGTAATGTTGAATATTGGATAAGGCTG
151 TGTGTCCTTTGTCTTGGGAGACAAAGCCCACAGCAGGTGGTGGTTGGGGT
GGTGGCAGCTCAGTGACAGGAGAGGTTTTTTTGCCTGTTTTTTTTTTTTT
TTTTTTTTTTAAGTAAGGTGTTCTTTTTTCTTAGTAAATTTTCTACTGGA
301 CTGTATGTTTTGACAGGTCAGAAACATTTCTTCAAAGAAGAACCCTTTTG
GAAACTGTACAGCCCTTTTCTTTTCATTCCCTTTTGTCTTCTGTGCCAAT
GCCTTTGGTTCTGATTGCATTATGGAACCGTTGATCGGAACCTTGAGGTT
451 TTTATTTATAGTGTGGCTTGAAAGCTTGGATAGCTGTTGTTACACGAGAT
ACCTTATTAAGTTTAGGCCAGCTTGATGCTTTATTTTTTCCCTTTGAAGT
AGTGAGCGTTCTCTGGTTTTTTTTTCCCTTTGAAACTGGTGAGGCTTAGATTT
601 TTCTAATGGGATTTTTTACCTGATGATCTAGTTGCATACCCAAATGCTTG
TAAATGTTTTCCCTAGTTAACATGTTGATAACTTCGGATTTACATGTTGTA
TATACTTGTCTCTGTGTTTCTAGTAAAAATATATGGCATTATATAGAAAT
751 ACGTAATTCCTGATTTCCCTTTTTTTTTATCTCTATGCTCTGTGTGTACAG
GTCAAACAGACTTCACCTCTATTTTTATTTATAGAATTTTATATGCAGTC
TGTCGTTGGTTCTTGTGTTGTAAGGATACAGCCTTAAATTTCCCTAGAGCG
901 ATGCTCAGTAAGGCGGGTGTGCATGGGTCAAATGTAAACGGGACAG
TTTGGCTGCTGCCTTCCCGAGATCCAGGACACTAAACTGCTTCTGCACCTG
AGGTATAAATCGCTTCAGATCCCAGGGAAGTGCAGATCCACGTGCATATT
1051 CTTAAAGAAGAATGAATACTTTCTAAATATTTTGGCATAGGAAGCAAGC
TGCATGGATTGTTTGGGACTTAAATTATTTGGTAACGGAGTGCATAGG
TTTTAAACACAGTTGCAGCATGCTAACGAGTCACAGCGTTTATGCAGAAG
1201 TGATGCCTGGATGCCTGTTGCAGCTGTTTACGGCACTGCCTTGCAGTGAG
CATTGCAGATAGGGGTGGGGTGCTTTGTGTCGTGTTCCACACGCTGCCA
CACAGCCACCTCCCGGAACACATCTCACCTGCTGGGTACTTTTCAAACCA
1351 TCTTAGCAGTAGTAGATGAGTTACTATGAAACAGAGAAGTTCCTCAGTTG
GATATTCTCATGGGATGTCTTTTTTCCCATGTTGGGCAAAGTATGATAAA
GCATCTCTATTTGTAAATTATGCACTTGTTAGTTCCTGAATCCTTTCTAT
1501 AGCACCACTTATTGCAGCAGGTGTAGGCTCTGGTGTGGCCTGTGTCTGTG
CTTCAATCTTTTAAAGCTTCTTTGGAAATACACTGACTTGATTGAAGTCT
CTTGAAGATAGTAAACAGTACTTACCTTTGATCCCAATGAAATCGAGCAT
1651 TTCAGTTGTAAAAGAATTCCGCCTATTTCATACCATGTAATGTAATTTTAC
ACCCCCAGTGCTGACACTTTGGAATATATTCAAGTAATAGACTTTGGCCT
CACCCTCTTGTGTACTGTATTTTGTAAATAGAAAATATTTTAAACTGTGCA
1801 TATGATTATTACATTATGAAAGAGACATTCTGCTGATCTTCAAATGTAAG
AAAATGAGGAGTGCGTGTGCTTTTATAAATACAAGTGATTGCAAATTAGT
GCAGGTGTCCTTAAAAAAAAAAAAAAAAAGTAATATAAAAAGGACCAGGT
1951 GTTTTACAAGTGAAATACATTCTATTTGGTAAACAGTTACATTTTATG
AAGATTACCAGCGCTGCTGACTTTCTAAACATAAGGCTGTATTGTCTTCC
TGTACCATTGCATTTCTCTATTCCCAATTTGCACAAGGATGTCTGGGTAA
2101 ACTATTCAAGAAATGGCTTTGAAATACAGCATGGGAGCTTGTCTGAGTTG
GAATGCAGAGTTGCACTGCAAAATGTCAGGAAATGGATGTCTCTCAGAAT
GCCCAACTCCAAAGGATTTTATATGTGTATATAGTAAGCAGTTTCTTGAT
2251 TCCAGCAGGCCAAAGAGTCTGCTGAATGTTGTGTTGCCGGAGACCTGTAT
TTCTCAACAAGGTAAGATGGTATCCTAGCAACTGCGGATTTTAAATACATT

TTCAGCAGAAGTACTTAGTTAATCTCTACCTTTAGGGATCGTTTCATCAT
2401 TTTTAGATGTTATACTTGAAATACTGCATAACTTTTAGCTTTCATGGGTT
CCTTTTTTTCAGCCTTTAGGAGACTGTTAAGCAATTTGCTGTCCAACCTTT
TGTGTTGGTCTTAAACTGCAATAGTAGTTTACCTTGTATTGAAGAAATAA
2551 AGACCATTTTTATATTAAAAAATACTTTTGTCTGTCTTCATTTTGACTTG
TCTGATATCCTTGAGTGCCCATATGTCAGTTCTGTCAGATATTCAGAC
ATCAAACTTAACGTGAGCTCAGTGGAGTTACAGCTGCGGTTTTGATGCT
2701 GTTATTATTTCTGAACTAGAAATGATGTTGTCTTCATCTGCTCATCAAA
CACTTCATGCAGAGTGTAAGGCTAGTGAGAAATGCATACATTTATTGATA
CTTTTTTAAAGTCAACTTTTTTATCAGATTTTTTTTTTTCATTTGGAAATA
2851 TTGTTTTCTAGACTGCATAGCTTCTGAATCTGAAATGCAGTCTGATTGGC
ATGAAGAAGCACAGCACTCTTCATCTTACTTAAACTTCATTTTGGAATGA
AGGAAGTTAAGCAAGGGCACAGGTCCATGAAATAGAGACAGTGCGCTCAG
3001 GAGAAAGTGAACCTGGATTTCTTTGGCTAGTGTTCTAAATCTGTAGTGAG
GAAAGTAACACCCGATTCTTTGAAAGGGCTCCAGCTTTAATGCTTCCAAA
TTGAAGGTGGCAGGCAACTTGGCCACTGGTTATTTACTGCATTATGTCTC
3151 AGTTTTCGAGCTAACCTGGCTTCTCCACTATTGAGCATGGACTATAGCCT
GGCTTCAGAGGCCAGGTGAAGGTTGGGATGGGTGGAAGGAGTGCTGGGCT
GTGGCTGGGGGGACTGTGGGGACTCCAAGCTGAGCTTGGGGTGGGCAGCA
3301 CAGGGAAAAGTGTTGGGTAACATTTTTTAAGTACTGTGTTGCAAACGTCTC
ATCTGCAAATACGTAGGGTGTGTACTCTCGAAGATTAACAGTGTGGGTTT
AGTAATATATGGATGAATTCACAGTGGAAGCATTCAAGGGTAGATCATCT
3451 AACGACACCAGATCATCAAGCTATGATTGGAAGCGGTATCAGAAGAGCGA
GGAAGGTAAGCAGTCTTCATATGTTTTCCCTCCACGTAAAGCAGTCTGGG
AAAGTAGCACCCCTTGAGCAGAGACAAGGAAATAATTCAGGAGCATGTGC
3601 TAGGAGAACTTTCTTGCTGAATTCTACTTGCAAGAGCTTTGATGCCTGGC
TTCTGGTGCTTCTGCAGCACCTGCAAGGCCAGAGCCTGTGGTGAGCTG
GAGGGAAAGATTCTGCTCAAGTCCAAGCTTCAGCAGGTCATTGTCTTTGC
3751 TTCTTCCCCCAGCACTGTGCAGCAGAGTGGAAGTATGTCGAAGCCTCCT
GTCCACTACCTGTTGCTGCAGGCAGACTGCTCTCAGAAAAAGAGAGCTAA
CTCTATGCCATAGTCTGAAGGTAAAATGGGTTTTAAAAAAGAAAAACACAA
3901 AGGCAAAACCGGCTGCCCCATGAGAAGAAAGCAGTGGTAAACATGGTAGA
AAAGGTGCAGAAGCCCCCAGGCAGTGTGACAGGCCCTCCTGCCACCTAG
AGGCGGGAACAAGCTTCCCTGCCTAGGGCTCTGCCGCGAAGTGCGTGTT
4051 TCTTTGGTGGGTTTTGTTTTGGCGTTTGGTTTTGAGATTTAGACACAAGGG
AAGCCTGAAAGGAGGTGTTGGGCACTATTTTGGTTTTGTAAAGCCTGTACT
TCAAATATATATTTTTGTGAGGGAGTGTAGCGAATTGGCCAATTTAAAATA
4201 AAGTTGCAAGAGATTGAAGGCTGAGTAGTTGAGAGGGTAACACGTTTAAT
GAGATCTTCTGAACTACTGCTTCTAAACACTTGTTTGAGTGGTGAGACC
TTGGATAGGTGAGTGCTCTTGTTACATGTCTGATGCACTTGCTTGTCTT
4351 TTCCATCCACATCCATGCATTCCACATCCACGATTTGTCACTTATCCCA
TATCTGTCATATCTGACATACCTGTCTCTTCGTCACTTGGTCAGAAGAAA
CAGATGTGATAATCCCCAGCCGCCCAAGTTTGAGAAGATGGCAGTTGCT
4501 TCTTTCCCTTTTTCTTGCTAAGTAAGGATTTTCTCCTGGCTTTGACACCT
CACGAAATAGTCTTCTGCCTTACATTCTGGGCATTATTTCAAATATCTT
TGGAGTGCGCTGCTCTCAAGTTTGTGTCTTCTTACTCTTAGAGTGAATGC
4651 TCTTAGAGTGAAAGAGAAGGAAGAGAAGATGTTGGCCGAGTTCTCTGAT
GAACACACCTCTGAATAATGGCCAAAGGTGGGTGGGTTTTCTCTGAGGAAC
GGGCAGCGTTTGCTTCTGAAAGCAAGGAGCTCTGCGGAGTTGCAGTTATT
4801 TTGCAACTGATGGTGGAAGTGGTGCTTAAAGCAGATTCCCTAGGTTCCCT

GCTACTTCTTTTCCTTCTTGGCAGTCAGTTTATTTCTGACAGACAAACAG
CCACCCCTGCTGAGGCTTAGAAAGTATGTGGCTCTGCCTGGGTGTGTTA
4951 CAGCTCTGCCCTGGTGAAAGGGGATTAAAACGGGCACCATTTCATCCCAA
CAGGATCCTCATTCATGGATCAAGCTGTAAGGAACCTGGGGCTCCAACCTC
AAAACATTAATTGGAGTACGAATGTAATTAATACTGCATTCTCGCATTC
5101 TAAGTCATTTAGTCTGGACTCTGCAGCATGTAGGTCGGCAGCTCCCACTT
TCTCAAAGACCACTGATGGAGGAGTAGTAAAAATGGAGACCGATTTCAGAA
CAACCAACGGAGTGTGCGCAAGAACTGATGGAAATAATGCATGAATTG
5251 TGTGGTGGACATTTTTTTTAAATACATAAACTACTTCAAATGAGGTCGGA
GAAGGTCAGTGTTTTATTAGCAGCCATAAAACCAGGTGAGCGAGTACCAT
TTTTCTCTACAAGAAAAACGATTCTGAGCTCTGCGTAAGTATAAGTTCTC
5401 CATAGCGGCTGAAGCTCCCCCTGGCTGCCTGCCATCTCAGCTGGAGTGC
AGTGCCATTTCTTGGGGTTTCTCTCACAGCAGTAATGGGACAATACTTC
ACAAAAATTCTTTCTTTTCTGTCTGTTGGGATCCCTACTGTGCCCTCCT
5551 GGTTTTACGTTACCCCTGACTGTTCCATTACGCGGTTTGGAAAGAGAAA
AAGAATTTGGAAATAAAACATGTCTACGTTATCACCTCCTCCAGCATTTT
GGTTTTTAATTATGTCAATAACTGGCTTAGATTTGGAAATGAGAGGGGGT
5701 TGGGTGTATTACCGAGGAACAAAGGAAGGCTTATATAAACTCAAGTCTTT
TATTTAGAGAACTGGCAAGCTGTCAAAAACAAAAAGGCCTTACCACCAAA
TTAAGTGAATAGCCGCTATAGCCAGCAGGGCCAGCACGAGGGATGGTGCA
5851 CTGCTGGCACTATGCCACGGCCTGCTTGTGACTCTGAGAGCAACTGCTTT
GGAATGACAGCACTTGGTGCAATTTCTTTGTTTCAGAATGCGTAGAGC
GTGTGCTTGGCGACAGTTTTTCTAGTTAGGCCACTTCTTTTTTCTCTC
6001 TCCTCATTCTCCTAAGCATGTCTCCATGCTGGTAATCCCAGTCAAGTGAA
CGTTCAAACAATGAATCCATCACTGTAGGATTCTCGTGGTGATCAAATCT
TTGTGTGAGGTCTATAAAATATGGAAGCTTATTTATTTTCTGTTCTTCCA
6151 TATCAGTCTTCTCTATGACAATTCACATCCACCACAGCAAATTAAGGTG
AAGGAGGCTGGTGGGATGAAGAGGGTCTTCTAGCTTTACGTTCTTCTCTG
CAAGGCCACAGGAAAATGCTGAGAGCTGTAGAATACAGCCTGGGGTAAGA
6301 AGTTCAGTCTCCTGCTGGGACAGCTAACCGCATCTTATAACCCCTTCTGA
GACTCATCTTAGGACCAAATAGGGTCTATCTGGGGTTTTTGTTCCTGCTG
TTCTCTCTGGAAGGCTATCTCACTATTTCACTGCTCCCACGGTTACAAAC
6451 CAAAGATACAGCCTGAATTTTTTCTAGGCCACATTACATAAATTTGACCT
GGTACCAATATTGTTCTCTATATAGTTATTTCTTCCCACTGTGTTTAA
CCCCTTAAGGCATTTCAGAACAACATAGAATCATAGAATGGTTTGGATTGGA
6601 AGGGGCCTTAAACATCATCCATTTCCAACCCTCTGCCATGGGCTGCTTGC
CACCCACTGGCTCAGGCTGCCCAGGGCCCCATCCAGCCTGGCCTTGAGCA
CCTCCAGGGATGGGGCACCCACAGCTTCTCTGGGCAGCCTGTGCCAACAC
6751 CTCACCACTCTCTGGGTAAAGAATTCTCTTTTAACATCTAATCTAAATCT
CTTCTCTTTTAGTTTAAAGCCATTCTCTTTTCCCGTTGCTATCTGTCC
AAGAAATGTGTATTGGTCTCCCTCCTGCTTATAAGCAGGAAGTACTGGAA
6901 GGCTGCAGTGAGGTCTCCCCACAGCCTTCTCTTCTCCAGGCTGAACAAGC
CCAGCTCCTTCAGCCTGTCTTCGTAGGAGATCATCTTAGTGGCCCTCCTC
TGGACCCATTCCAACAGTTCCACGGCTTTCTTGTGGAGCCCCAGGTCTGG
7051 ATGCAGTACTTCAGATGGGGCCTTACAAAGGCAGAGCAGATGGGGACAAT
CGTTACCCCTCCCTGCTGGCTGCCCCTGTTTTGATGCAGCCAGGGTAC
TGTGGCCCTTTCAGGCTCCCAGACCCCTTGCTGATTTGTGTCAAGCTTTT
7201 CATCCACCAGAACCCACGCTTCCTGGTTAATACTTCTGCCCTCACTCTG
TAAGCTTGTTCAGGAGACTTCCATTCTTTAGGACAGACTGTGTTACACC
TACCTGCCCTATTCTTGCAATATATACATTTTCAGTTCATGTTTCTGTAAAC

7351 AGGACAGAATATGTATTCTCTAACAAAAATACATGCAGAATTCCTAGTG
CCATCTCAGTAGGGTTTTTCATGGCAGTATTAGCACATAGTCAATTTGCTG
CAAGTACCTTCCAAGCTGCGGCTCCCATAAATCCTGTATTTGGGATCAG
7501 TTACCTTTTGGGTAAGCTTTTGTATCTGCAGAGACCCTGGGGTTCTGA
TGTGCTTCAGCTCTGCTCTGTTCTGACTGCACCATTTTCTAGATCACCCA
GTTGTTCCCTGTACAACCTTCCTTGTCTCCATCCTTTCCCAGCTTGTATCT
7651 TTGACAAATACAGGCCTATTTTTGTGTTTGTCTTCAGCAGCCATTTAATTC
TTCAGTGTCTCTTGTCTGTTGATGCCACTGGAACAGGATTTTCAGCAG
TCTTGCAAAGAACATCTAGCTGAAAACCTTCTGCCATTCAATATTTCTTAC
7801 CAGTTCTTCTTGTGTTGAGGTGAGCCATAAATTACTAGAACTTCGTCAGT
ACAAGTTTTATGCATTTTATTACTTCTATTATGTACTTACTTTGACATAAC
ACAGACACGCACATATTTTGTGCTGGGATTTCCACAGTGTCTCTGTGTCTT
7951 CACATGGTTTTACTGTCATACTTCCGTTATAACCTTGGCAATCTGCCCAG
CTGCCCATCACAAAGAAAAGAGATTCTTTTTTTATTACTTCTCTTCAGCCA
ATAAACAAAATGTGAGAAGCCCAACAAGAACTTGTGGGGCAGGCTGCCA
8101 TCAAGGGAGAGACAGCTGAAGGGTTGTGTAGCTCAATAGAATTAAGAAAT
AATAAAGCTGTGTCTCAGACAGTTTTGCCTGATTTATACAGGCACGCCCAA
GCCAGAGAGGGCTGTCTGCCAAGGCCACCTTGCACTCCTTGGTTTTGTAAGA
8251 TAAGTCATAGGTAACCTTTTCTGGTGAATTGCGTGGAGAATCATGATGGCA
GTTCTTGCTGTTTACTATGGTAAGATGCTAAAAATAGGAGACAGCAAAGTA
ACACTTGCTGCTGTAGGTGCTCTGCTATCCAGACAGCGATGGCACTCGCA
8401 CACCAAGATGAGGGATGCTCCCAGCTGACGGATGCTGGGGCAGTAACAGT
GGGTCCCATGCTGCCTGCTCATTAGCATCACCTCAGCCCTCACCAGCCCA
TCAGAAGGATCATCCCAAGCTGAGGAAAGTTGCTCATCTTCTTCACATCA
8551 TCAAACCTTTGGCTGACTGATGCCTCCCGGATGCTTAAATGTGGTCACT
GACATCTTTATTTTTCTATGATTTCAAGTCAGAACCTCCGGATCAGGAGG
GAACACATAGTGGGAATGTACCCTCAGCTCCAAGGCCAGATCTTCTTCA
8701 ATGATCATGCATGCTACTTAGGAAGGTGTGTGTGTGTGAATGTAGAATTG
CCTTTGTTATTTTTCTTCTGCTGTCAGGAACATTTTGAATACCAGAGA
AAAAGAAAAGTGCTCTTCTTGGCATGGGAGGAGTTGTCACTTGC AAAA
8851 TAAAGGATGCAGTCCCAAATGTTTATAATCTCAGGGTCTGAAGGAGGATC
AGAACTGTGTATACAATTTAGGCTTCTCTGAATGCAGCTTTTGAAAGC
TGTTCCTGGCCGAGGCAGTACTAGTCAGAACCTCGGAAACAGGAACAAA
9001 TGTCTTCAAGGTGCAGCAGGAGGAAACACCTTGCCCATCATGAAAGTGAA
TAACCACTGCCGCTGAAGGAATCCAGCTCCTGTTTGAGCAGGTGCTGCAC
ACTCCCACACTGAAACAACAGTTTATTTTTATAGGACTTCCAGGAAGGAT
9151 CTTCTTCTTAAGCTTCTTAATTATGGTACATCTCCAGTTGGCAGATGACT
ATGACTACTGACAGGAGAATGAGGAAGTACTGGGAATATTTCTGTTTGA
CCACCATGGAGTCACCCATTTCTTTACTGGTATTTGGAAATAATAATCT
9301 GAATTGCAAAGCAGGAGTTAGCGAAGATCTTCATTTCTTCCATGTTGGTG
ACAGCACAGTTCTGGCTATGAAAGTCTGCTTACAAGGAAGAGGATAAAAA
9401 TCATAGGGATAATAAATCTAAGTTTGAAGACAATGAGGTTTTAGCTGCAT
TTGACATGAAGAAATTGAGACCTCTACTGGATAGCTATGGTATTTACGTG
TCTTTTTGCTTAGTTACTTATTGACCCAGCTGAGGTCAAGTATGAACTC
9551 AGGTCTCTCGGGCTACTGGCATGGATTGATTACATACAACTGTAATTTA
GCAGTGATTTAGGGTTTATGAGTACTTTTGCAGTAAATCATAGGGTTAGT
AATGTTAATCTCAGGGAAAAAAGCCAACCCTGACAGACATCCC
9701 AGCTCAGGTGGAAATCAAGGATCACAGCTCAGTGCGGTCCCAGAGAACAC
AGGGAATCTTCTTAGGACCTTTATGTACAGGGCCTCAAGATAACTGAT
GTTAGTCAGAAGACTTTCATTCTGGCCACAGTTCAGCTGAGGCAATCCT

9851 GGAATTTTCTCTCCGCTGCACAGTTCCAGTCATCCCAGTTTGTACAGTTC
TGGCACTTTTTGGGTCAGGCCGTGATCCAAGGAGCAGAAGTTCCAGCTAT
GGTCAGGGAGTGCCTGACCGTCCCAACTCACTGCACTCAAACAAAGGCGA
10001 AACCACAAGAGTGGCTTTTGTGAAATTGCAGTGTGGCCCAGAGGGGCTG
CACCAGTACTGGATTGACCACGAGGCAACATTAATCCTCAGCAAGTGCAA
TTTGCAGCCATTAAATTGAACTAACTGATACTACAATGCAATCAGTATCA
10151 ACAAGTGGTTTGGCTTGGAGATGGAGTCTAGGGGCTCTACAGGAGTAGC
TACTCTCTAATGGAGTTGCATTTTGAAGCAGGACACTGTGAAAAGCTGGC
CTCCTAAAGAGGCTGCTAAACATTAGGGTCAATTTTCCAGTGCACCTTCT
10301 GAAGTGTCTGCAGTTCCCATGCAAAGCTGCCCCAAACATAGCACTTCCAA
TTGAATACAATTATATGCAGGCGTACTGCTTCTTGCCAGCACTGTCCCTTC
TCAAATGAACTCAACAAACAATTTCAAAGTCTAGTAGAAAGTAACAAGCT
10451 TTGAATGTCATTAAAAAGTATATCTGCTTTCAGTAGTTCAGCTTATTTAT
GCCCCACTAGAAACATCTTGTACAAGCTGAACACTGGGGCTCCAGATTAGT
GGTAAAACCTACTTTATACAATCATAGAATCATAGAATGGCCTGGGTTGG
10601 AAGGGACCCCCAAGGATCATGAAGATCCAACACCCCCGCCACAGGCAGGGC
CACCAACCTCCAGATCTGGTACTAGACCAGGCAGCCCAGGGCTCCATCCA
ACCTGGCCATGAACACCTCCAGGGATGGAGCATCCACAACCTCTCTGGGC
10751 AGCCTGTGCCAGCACCTCACCACCCTCTCTGTGAAGAACTTTCCCTGAC
ATCCAATCTAAGCCTTCCCTCCTTGAGGTTAGATCCACTCCCCCTTGTGC
TATCACTGTCTACTCTTGTAAAAAGTTGATTCTCCTCCTTTTTTGAAGGT
10901 TGCAATGAGGTCTCCTTGCAGCCTTCTTCTCTTCTGCAGGATGAACAAGC
CCAGCTCCCTCAGCCTGTCTTTATAGGAGAGGTGCTCCAGCCCTCTGATC
ATCTTTGTGGCCCTCCTCTGGACCCGCTCCAAGAGCTCCACATCTTTCCT
11051 GTACTGGGGGCCCCAGGCCTGAATGCAGTACTCCAGATGGGGCCTCAAAA
GAGCAGAGTAAAGAGGGACAATCACCTTCCCTCACCTGCTGGCCAGCCCT
CTTCTGATGGAGCCCTGGATACAACCTGGCTTTCTGAGCTGCAACTTCTCC
11201 TTATCAGTTCCACTATTAAAACAGGAACAATAACAACAGGTGCTGATGGCC
AGTGCAGAGTTTTTTCACACTTCTTCATTTCCGGTAGATCTTAGATGAGGAA
CGTTGAAGTTGTGCTTCTGCGTGTGCTTCTTCCCTCCTCAAATACTCCTGC
11351 CTGATACCTCACCCACCTGCCACTGAATGGCTCCATGGCCCCCTGCAGC
CAGGGCCCTGATGAACCCGGCACTGCTTCAGATGCTGTTTAATAGCACAG
TATGACCAAGTTGCACCTATGAATACACAAACAATGTGTTGCATCCTTCA
11501 GCACCTTGAGAAGAAGAGCCAAATTTGCATTGTGAGGAAATGGTTTAGTAA
TTCTGCCAATTAAAACCTTGTTTATCTACCATGGCTGTTTTTATGGCTGTT
AGTAGTGGTACACTGATGATGAACAATGGCTATGCAGTAAAATCAAGACT
11651 GTAGATATTGCAACAGACTATAAAATTCCTCTGTGGCTTAGCCAATGTGG
TACTTCCCACATTGTATAAGAAATTTGGCAAGTTTAGAGCAATGTTTGAA
GTGTTGGGAAATTTCTGTATACTCAAGAGGGCGTTTTTGACAACTGTAGA
11801 ACAGAGGAATCAAAAGGGGGTGGGAGGAAGTTAAAAGAAGAGGCAGGTGC
AAGAGAGCTTGCAGTCCCGCTGTGTGTACGACACTGGCAACATGAGGTCT
TTGCTAATCTTGGTGCTTTGCTTCCTGCCCCCTGGCTGCCTTAGGGTGCGA
11951 TCTGCCTCAGACCCACAGCCTGGGCAGCAGGAGGACCCCTGATGCTGCTGG
CTCAGATGAGGAGAATCAGCCTGTTTAGCTGCCTGAAGGATAGGCACGAT
TTTGGCTTTCCTCAAGAGGAGTTTGGCAACCAGTTTCAGAAGGCTGAGAC
12101 CATCCCTGTGCTGCACGAGATGATCCAGCAGATCTTTAACCTGTTTAGCA
CCAAGGATAGCAGCGCTGCTTGGGATGAGACCTGCTGGATAAGTTTTTAC
ACCGAGCTGTACCAGCAGCTGAACGATCTGGAGGCTTGCCTGATCCAGGG
12251 CGTGGGCGTGACCGAGACCCCTCTGATGAAGGAGGATAGCATCCTGGCTG
TGAGGAAGTACTTTCAGAGGATCACCTGTACCTGAAGGAGAAGAAGTAC

AGCCCCCTGCGCTTGGGAAGTCGTGAGGGCTGAGATCATGAGGAGCTTTAG
12401 CCTGAGCACCAACCTGCAAGAGAGCTTGAAGTCTAAGGAGTAAAAAGTCT
AGAGTCGGGGCGGCGCGTGGTAGGTGGCGGGGGTTCCCAGGAGAGCCCC
CAGCGCGGACGGCAGCGCCGTCACCTACCGCTCCGTCTCCCTCCGCCCAG
12551 GGTGCGCTGGCGCAACCGCTGCAAGGGCACCGACGTCCAGGCGTGGATCA
GAGGCTGCCGGCTGTGAGGAGCTGCCGCGCCCCGGCCCCGCTGCACAG
CCGGCCGCTTTGCGAGCGCGACGCTACCCGCTTGGCAGTTTAAACGCAT
12701 CCCTCATTAAACGACTATACGCAAACGCCTTCCCGTCGGTCCGCGTCTC
TTTCCGCCGCCAGGGCGACACTCGCGGGGAGGGCGGGAAGGGGGCCGGGC
GGGAGCCCCGCGGCCAACCGTCGCCCCGTGACGGCACCGCCCCGCCCCCGT
12851 GACGCGGTGCGGGCGCCGGGGCCGTGGGGCTGAGCGCTGCGGCGGGGCCG
GGCGGGGCCGGGGCGGAGCTGAGCGCGCGCGGCTGCGGGCGGCGCCCC
CTCCGGTGCAATATGTTCAAGAGAATGGCTGAGTTCGGGCCTGACTCCGG
13001 GGGCAGGGTGAAGGTGCGGCGCGGGCGGAGGGACGGGGCGGGCGCGGGGC
CGCCCCGGCGGGTGCCGGGGCCCTCTGCCGGCCCCGCCGGCTCGGGCTGCTG
CGGCGCTTACGGGCGCGCTTCTCGCCGCTGCCGCTTCTTCTCTCCCCG
13151 GCAAGGGCGTCAACATCGTGAAGCCGGTAGTGTACGGGAACGTGGCGCGG
TACTTCGGGAAGAAGAGGGAGGAGGACGGGCACACGCATCAGTGGACGGT
TTACGTGAAGCCCTACAGGAACGAGGTAGGGCCCCGAGCGCGTCCGGCCG
13301 GTTCTCGGAGCGCCGGAGCCGTCAGCGCCGCGCCTGGGTGCGCTGTGGGA
CACAGCGAGCTTCTCTCGTAGGACATGTCCGCCTACGTGAAAAAATCCA
GTTCAAGCTGCACGAGAGCTACGGGAATCCTCTCCGAGGTGGGTGTTGCG
13451 TCGGGGGGTTTGCTCCGCTCGGTCCCGCTGAGGCTCGTCGCCCTCATCTT
TCTTTCGTGCCGCAGTCGTTACCAAACCGCCGTACGAGATCACCGAAACG
13551 GGCTGGGGCGAATTTGAAATCATCATCAAGATATTTTTCATTGATCCAAA
CGAGCGACCCGTAAGTACGCTCAGCTTCTCGTAGTGCTTCCCCCGTCTTG
GCGGCCCGGGCTGGGCTGCTCGCTGCTGCCGGTCACAGTCCCGCCAGCC
13701 GCGGAGCTGACTGAGCTCCCTTTCCCGGGACGTGTGCTCTGTGTTCCGTC
AGCGAGGCTATCGGGAGGGCTTTGGCTGCATTTGGCTTCTCTGGCGCTTA
GCGCAGGAGCACGTTGTGCTACGCCTGAACTACAGCTGTGAGAAGGCCGT
13851 GGAAACCGCTCTCAAACTGATTTATTGGCGAAATGGCTCTAAACTAAATC
GTCTCCTCTCTTTGGAAAATGCTTTAGAGAAGGTCTCTGTGGTAGTTCTTA
TGCATCTATCCTAAAGCACTTGGCCAGACAATTTAAAGACATCAAGCAGC
14001 ATTTATAGCAGGCACGTTTAATAACGAATACTGAATTTAAGTAACTCTGC
TCACGTTGTATGACGTTTATTTTCGTATTCCTGAAAGCCATTAAAATCCT
GTGCAGTTGTTTAGTAAGAACAGCTGCCACTGTTTTGTATCTAGGAGATA
14151 ACTGGTGTTTTCCCTACAGTTTCAAGCTGATAAACTCTGTCTTTGTATC
TAGGTAACCCGTGTATCACTTGCTGAAGCTTTTTTCAGTCTGACACCAATGC
AATCCGCGGAAAGAAAACGTAGTTTCTGAATTCTATGATGAAATGGTAT
14301 GAAAATTTTAATGTCAACCGAGCCTGACTTTATTTAAAAAAAATTATTGA
TGGTGCTGTGTATTTTGGTCCTTCCTTAGATATTTCAAGATCCTACTGCC
ATGATGCAGCAACTGCTAACGACGTCCCGTCAGCTGACACTTGGTGCTTA
14451 CAAGCATGAAACAGAGTGTAAGTGCAAATGAGGATACCTTCGCCGACCG
TCATTCACACTAATGTTTTCTGTGGGATGTGATCGTACAGTGAGTTTGG
CTGTGTGAAATTTGAATAGCTTGGTATTGGCAGTGATGACGTGATCGATG
14601 CCTTGCTTATCATGTTTGAATGAAGTAGAATAAATGCAGCCTGCTTTAT
TTGAGATAGTTTGGTTCATTTTATGGAATGCAAGCAAAGATTATACTTCC
TCACTGAATTGCACTGTCCAAAGGTGTGAAATGTGTGGGGATCTGGAGGA
14751 CCGTGACCGAGGGACATTGGATCGCTATCTCCCATTTCTTTTGCTGTTAC
CAGTTCAGATTTTCTTTTCACCTAGTCTTTAATTCCCAGGGTTTTGTTTT

TTCTTGGTCATAGTTTTTGTCTTTTCACTCTGGCAAATGATGTTGTGAAT
14901 TACACTGCTTCAGCCACAAAAGCTGATGGACTGAATGAGGTCATCAAACAA
ACTTTTCTTCTTCCGTATTTCTTTTTTTTCCCCCACTTATCATTTTTTAC
TGCTGTTGTTGAGTCTGTAAGGCTAAAAGTAAGTGTGTTTGTGCTTTTTTCA
15051 GGACGTGTGCTTTCCAAATTACTGCCACATATATAAAGAAAGGTTGGAAT
TTTAAAGATAATTCATGTTTCTTCTTTTTTTGCCACCACAGTTGCAGA
TCTTGAAGTAAAAACCAGGGAAAAGCTGGAAGCTGCCAAAAAGAAAACCA
15201 GTTTTGAAATGCTGAGCTTAAAGAAAGGTTAAAGCAAGTCGTGAAACC
ATCAACTGCTTAAAGAGTGAAATCAGAAAAGCTCGAAGAGGATGATCAGTC
TAAAGATATGTGATGAGTGTGACTTGGCAGGGAGCCTATAATGAGAATG
15351 AAAGGACTTCAGTCGTGGAGTTGTATGCGTTCTCTCCAATTCTGTAACGG
AGACTGTATGAATTTTCATTTGCAAATCACTGCAGTGTGTGACAACTGACT
TTTTATAAATGGCAGAAAACAAGAATGAATGTATCCTCATTTTATAGTTA
15501 AAATCTATGGGTATGTACTGGTTTATTTCAAGGAGAATGGATCGTAGAGA
CTTGGAGGCCAGATTGCTGCTTGTATTGACTGCATTTGAGTGGTGTAGGA
ACATTTTGTCTATGGTCCCGTGTAGTTTACAGAATGCCACTGTTCACTG
15651 TTTTGTGTTTGTATTTTACTTTTTCTACTGCAACGTCAAGGTTTTAAAGT
TGAAAATAAAACATGCAGGTTTTTTTTTAAATATTTTTTTGTCTCTATCCA
GTTTGGGCTTCAAGTATTATTGTTAACAGCAAGTCCTGATTTAAGTCAGA
15801 GGCTGAAGTGTAATGGTATTCAAGATGCTTAAGTCTGTTGTCAGCAAAC
AAAAGAGAAAACCTCATAAAATCAGGAAGTTGGCATTCTCTAATAACTTCT
TTATCAACAGATAAGAGTTTCTAGCCCTGCATCTACTTTCACTTATGTAG
15951 TTGATGCCTTTATATTTTGTGTGTTTGGATGCAGGAAGTGATTCCTACTC
TGTTATGTAGATATTTCTATTTAACACTTGTAAGTCTGCTGTGCTTAGCCTT
16051 TCCCCATGAAAAATTCAGCGGCTGTAAATCCCCCTCTTCTTTTGTAGCCTC
ATACAGATGGCAGACCCTCAGGCTTATAAAGGCTTGGGCATCTTCTTTAC
TGCTTTGAGATTCTGTGTTGCAGTAACCTCTGCCAGAGAGGAGAAAAGCC
16201 CCACAAACCTCATCCCCCTTCTTCTATAGCAATCAGTATTACTAATGCTTT
GAGAACAGAGCACTGGTTTGAACGTTTGATAATTAGCATTTAACATGGC
TTGGTAAAGATGCAGAACTGAAACAGCTGTGACAGTATGAACTCAGTATG
16351 GAGACTTCATTAAGACAAACAGCTGTTAAATCAGGCATGTTTCATTGAG
GAGGACGGGGCAACTTGCACCAAGTGGTGCCACACAAATCCTTCTGGCG
CTGCAGACCAATTTTTCTGGCATTCTGACTGCCGTTGCTGCTGGTCACAG
16501 AGAGCAACTATTTTTATCAGCCACAGGCAATTTGCTTGTAGTATTTCCA
AGTGTGTTAGGTAAGTATAAATGCATCGGCTCCAGAGCACTTTGAGTATA
CTTATTAATAAACATAAATGAAAGACAAATTAGCTTTGCTTGGGTGCACAG
16651 AACATTTTTAGTTCCAGCCTGCTTTTTTGGTAGAAGCCCTCTTCTGAGGCT
AGAAGTGAAGTTGACAAGTAGAGAAAGTGGCAACGGAGCTATTGCTATCG
16751 AAGGATCCTTGTTAACAAAGTTAATCGTCTTTTAAGGTTTGGTTTATTCA
TTAAATTTGCTTTTAAGCTGTAGCTGAAAAAGAACGTGCTGTCTCCATG
16851 CACCAGGTGGCAGCTCTGTGCAAAGTGCTCTCTGGTCTCACCAGCCTTTT
AATTGCCGGGATTCTGGCACGTCTGAGAGGGCTCAGACTGGCTTCGTTTG
TTTGAACAGCGTGTAAGTCTTCTGTAGACATGGCCGGTTTCTCTCCTGC
17001 AGCTTATGAAACTGTTTCACTGAACACACTGGAACAGGTTGCCCAAGGA
GGCCGTGGATGCCCCATCCCTGGAGGCATTCAAGGCCAGGCTGGATGTGG
CTCTGGGCAGCCTGGTCTGGTGGTGGCGATCTGCACATAGCAGCGGGG
17151 TTGAAACTCGATGATCACTGTGGTCCTTTTCAACCCAGGCTATTCTATGA
TTCTATGATTCAACAGCAAATCATATGTACTGAGAGAGGAAACAAACACA
AGTGCTACTGTTTGCAAGTTTTGTTTCATTTGGTAAAAGAGTCAGGTTTTA
17301 AAATTCAAATCTGTCTGGTTTTTGGTGTTTTTTTTTTTTATTTATTATT

TCTTTGGGGTTCTTTTTGATGCTTTATCTTTCTCTGCCAGGACTGTGTGA
CAATGGGAACGAAAAAGAACATGCCAGGCACTGTCCTGGATTGCACACGC
17451 TGGTTGCACTCAGTAGCAGGCTCAGAACTGCCAGTCTTTCCACAGTATTA
CTTTCTAAACCTAATTTTAATAGCGTTAGTAGACTTCCATCACTGGGCAG
TGCTTAGTGAATGCTCTGTGTGAACGTTTTACTTATAAGCATGTTGGAAG
17601 TTTTGATGTTCTGGATGCAGTAGGGAAGGACAGATTAGCTATGTGAAAA
GTAGATTCTGAGTATCGGGGTACAAAAAGTATAGAAACGATGAGAAATT
CTTGTTGTAACATAATTGGAATTTCTTTAAGCGTTCACCTATGCTACATTC
17751 ATAGTATTTCCATTTAAAAGTAGGAAAAGGTAAAACGTGAAATCGTGTGA
TTTTCGGATGGAACACCGCCTTCCTATGCACCTGACCAACTTCCAGAGGA
AAAGCCTATTGAAAGCCGAGATTAAGCCACCAAAAAGAACTCATTTGCATT
17901 GGAATATGTAGTATTTGCCCTCTTCCTCCCGGGTAATTACTATACTTTAT
AGGGTGCTTATATGTTAAATGAGTGGCTGGCACTTTTTATTCTCACAGCT
GTGGGGAATTCTGTCCTCTAGGACAGAAACAATTTTAATCTGTTCCACTG
18051 GTGACTGCTTTGTCAGCACTTCCACCTGAAGAGATCAATACACTCTTCAA
TGTCTAGTTCTGCAACACTTGGCAAACCTCACATCTTATTTTCACTCTC
TTCATGCCTATGCTTATTAAAGCAATAATCTGGGTAATTTTTGTTTTAAT
18201 CACTGTCTCTGACCCAGTGATGACCGTGTCCACCTAAAGCTCAATTCAG
GTCCTGAATCTCTTCAACTCTCTATAGCTAACATGAAGAATCTTCAAAAG
TTAGGTCTGAGGGACTTAAGGCTAACTGTAGATGTTGTTGCCTGGTTTTCT
18351 GTGCTGAAGGCCGTGTAGTAGTTAGAGCATTC AACCTCTAGAAGAAGCTT
GGCCAGCTGGTCGACCTGCAGATCCGGCCCTCGAGGGGGGGCCCGGTACC
CAGCTTTTTGTTCCCTTTAGTGAGGGTTAATTTTCGAGCTTGGCGTAATCAT
18501 GGTATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCACAC
AACATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCTTAATGAGT
GAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCCGCTTTCCAGTCGG
18651 GAAACCTGTGCTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA
GGCGGTTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCT
GCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGG
18801 TAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGA
GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC
GTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCT
18951 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT
CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTAC
CGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATA
19101 GCTCACGCTGTAGGTATCTCAGTTTCGGTGTAGGTCGTTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTTCAGCCGACCGCTGCGCCTTATCCGG
TAATATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
19251 CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT
ACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGGACAGT
ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTG
19401 GTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTT
GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACCTCACGTT
19551 AAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTT
TTAAATTAAAAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAAC
TTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGA
19701 TCTGTCTATTTTCGTTTATCCATAGTTGCCTGACTCCCCGTCGTGTAGATA
ACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACC
GCGAGACCCACGCTCACC GGCTCCAGATTTATCAGCAATAAACCAGCCAG

19851 CCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATC
CAGTCTATTAATTGTTGCCGGAAGCTAGAGTAAGTAGTTCGCCAGTTAA
TAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACAGCT
20001 CGTCGTTTGGTATGGCTTCATTAGCTCCGGTTCCCAACGATCAAGGCGA
GTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCC
TCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTA
20151 TGGCAGCACTGCATAATTCTCTTACTGTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCG
GCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCAC
20301 ATAGCAGAACTTTAAAAGTGCTCATCATTTGAAAAACGTTCTTCGGGGCGA
AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCAC
TCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTG
20451 GGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG
ACACGGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTT
20601 AGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCA
CCTAAATTGTAAGCGTTAATATTTTGTAAAATTCGCGTTAAATTTTGT
TAAATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTA
20751 TAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGA
ACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAA
ACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCTAATCAAG
20901 TTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGA
GCCCCGATTTAGAGCTTGACGGGGAAGCCGGCGAACGTGGCGAGAAAG
GAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGC
21051 GGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTAC
AGGGCGCGTCCCATTTCGCCATTAGGCTGCGCAACTGTTGGGAAGGGCGA
TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGC
21201 TGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTCCCAGTCACGACGTT
GTAAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGGCGAATTG
21301 GAGCTCCACCGCGGTGGCGGCCGCTCTAG

Fig. 8

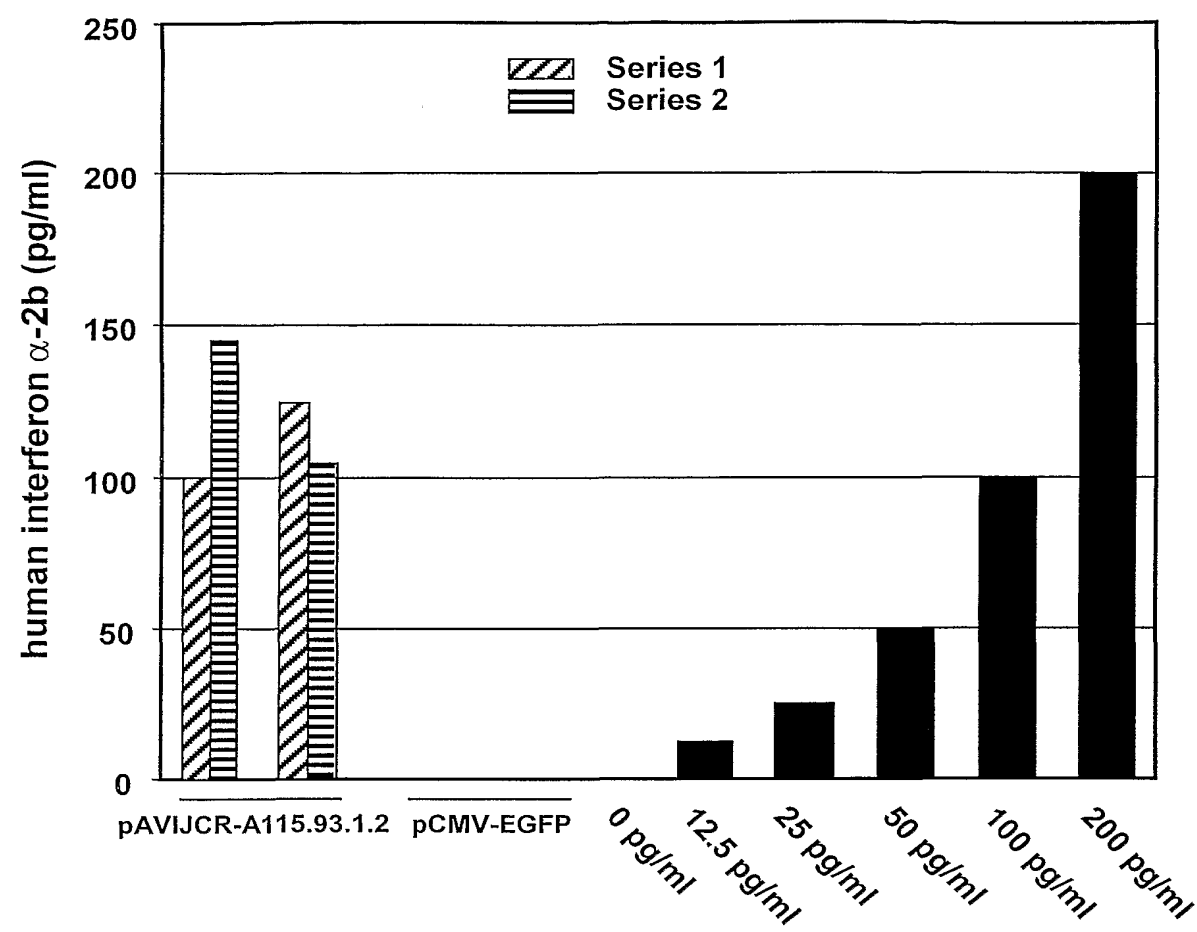
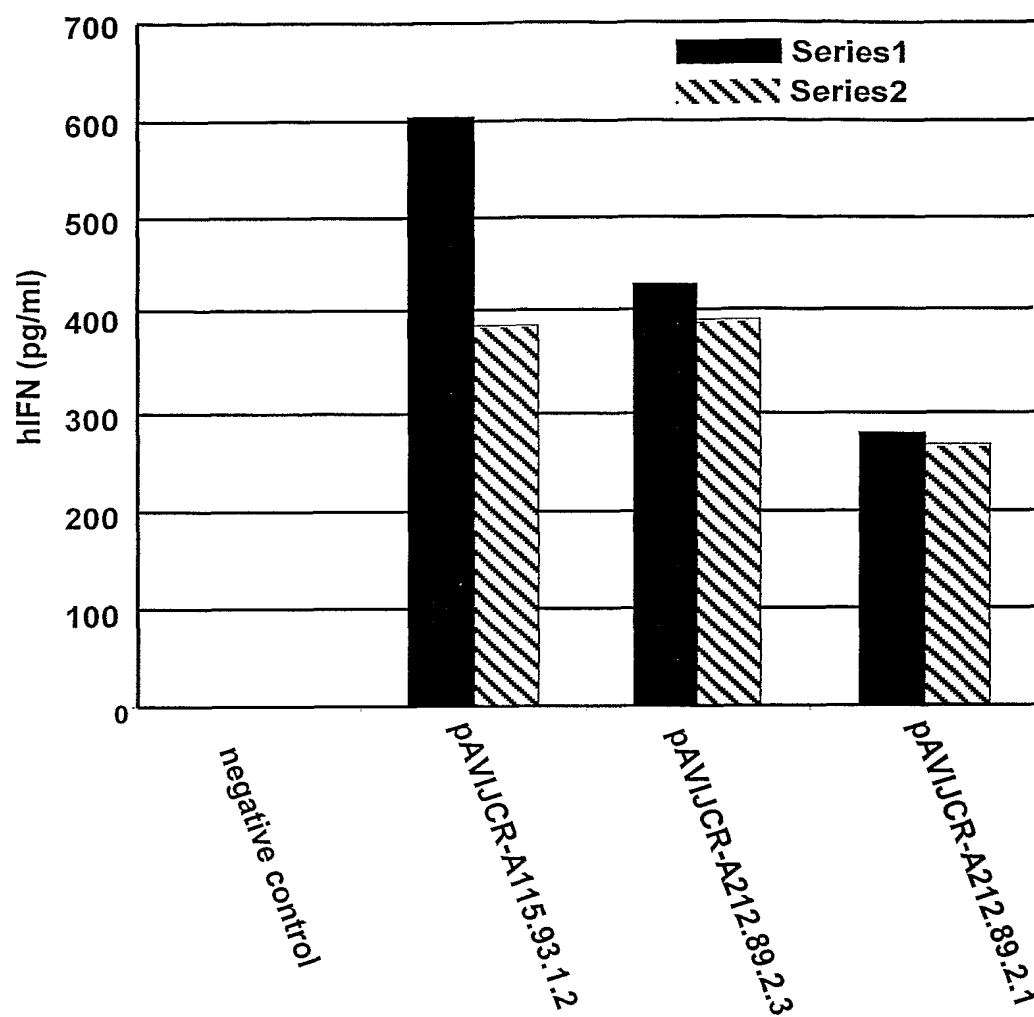


Fig. 9

**Fig. 10**

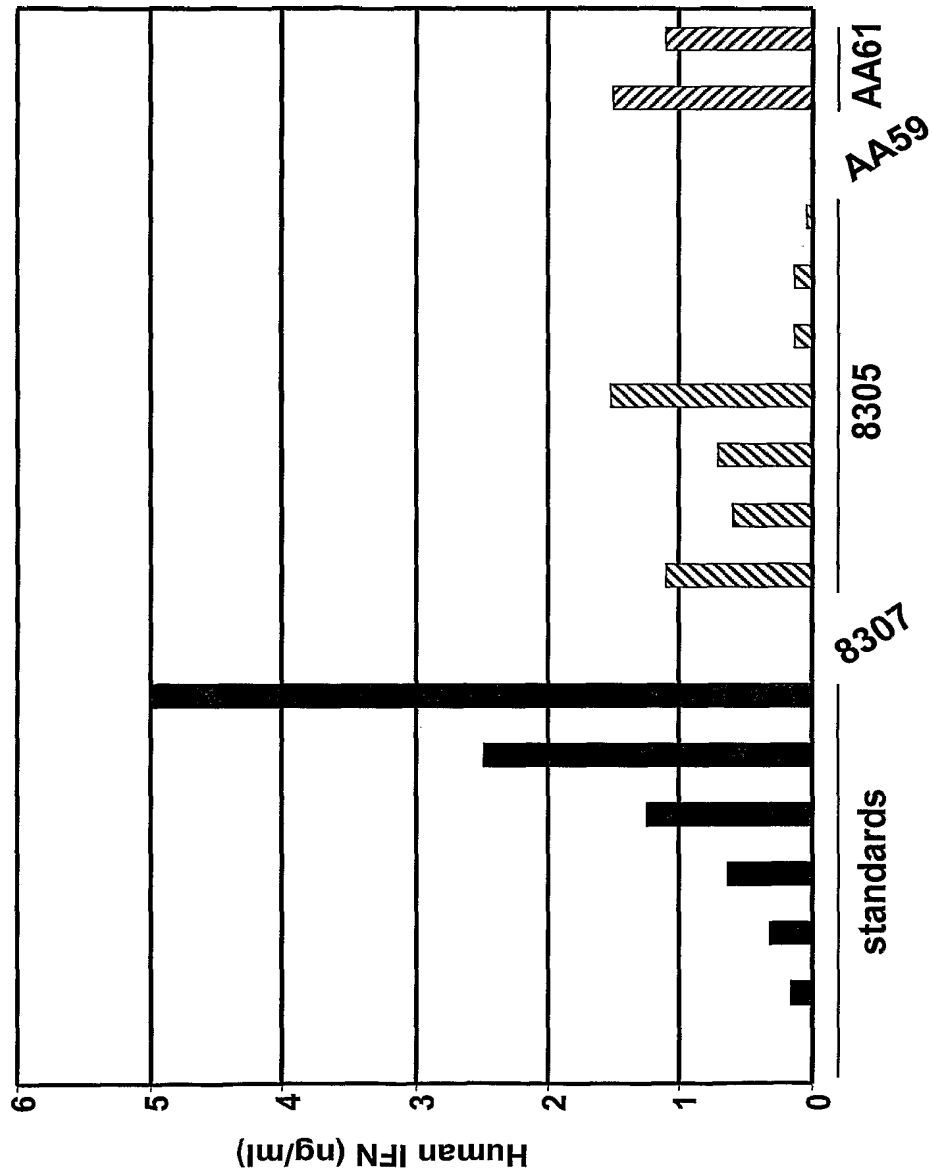


Fig. 11

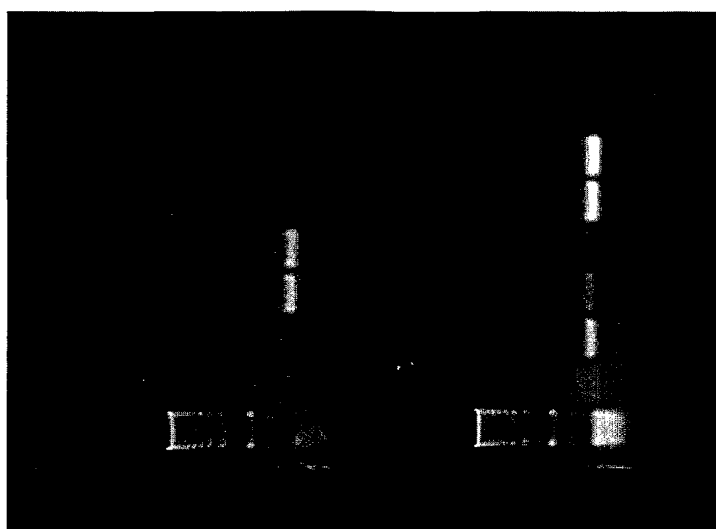


Fig. 12